

Thesis for doctoral degree (Ph.D.)
2010

ACQUISITION AND MANIPULATION OF MENTAL STRUCTURES:

Investigations on artificial grammar
learning and implicit sequence
processing

Christian Forkstam

Thesis for doctoral degree (Ph.D.) 2010

ACQUISITION AND MANIPULATION OF MENTAL STRUCTURES

Christian Forkstam



**Karolinska
Institutet**

200
1810 – 2010 *Years*



**Karolinska
Institutet**

200
1810 – 2010 *Years*

From THE DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

**ACQUISITION AND
MANIPULATION OF MENTAL
STRUCTURES:**

**Investigations on artificial
grammar learning and
implicit sequence
processing**

Christian Forkstam



**Karolinska
Institutet**

Stockholm 2010

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by Larserics AB.

© Christian Forkstam, 2010
ISBN 978-91-7409-769-6

Till en viss förvåning

ABSTRACT

This thesis introduces repetitive artificial grammar learning as a paradigm in the investigation of sequential implicit learning, in particular as a model for language acquisition and processing. Implicit learning of sequential structure captures an essential cognitive processing capacity of interest from a larger cognitive neuroscience perspective. We investigate in this thesis the underlying neural processing architecture for implicit learning/acquisition to acquire and process non-motor sequences, an implicit non-motor procedural learning ability present in the human cognitive system. In doing this, we validate and explore the repeated artificial grammar learning paradigm as a laboratory model to investigate the acquisition and processing of structural aspects of language, e.g. (morpho-) syntax processing, to further our understanding of the specific neural processing architecture subserving the syntax processing ability of the language faculty. A theoretical background on sequential procedural learning and formal grammars in cognitive processing is presented together with a general outline of the neuronal implementation of the cognitive functions involved. We suggest a lexical view on the processing and acquisition of artificial grammars to be beneficial to understand the nature and representation of the acquired knowledge. From this perspective we suggest that formal grammar acquisition and processing of the (regular) grammar type commonly studies in artificial grammar learning can be used as a model to investigate the neuronal infrastructure supporting language acquisition and processing, including to characterize the neuronal infrastructure supporting syntax processing and unification (cf. e.g., Hagoort, 2003; Jackendoff, 1997; Jackendoff, 2007; Kaan & Swaab, 2002; Shieber, 1986; Vosse & Kempen, 2000).

In **study 1** we describe the neuronal implementation using a setup based on the seminal study on implicit learning by Reber (1967), and report an overlap in the neural activation on artificial syntax violation and similar natural syntax violation. In **study 2** we replicate this finding using a more elaborated model with repeated acquisition sessions to simulate a prolonged acquisition period, and using a sequential presentation forcing the cognitive processing into a sequential processing mode. A neuronal activation pattern is reported which suggests that frontostriatal circuits are at play during artificial grammar classification, specifically the left inferior frontal region Brodmann's area 44/45 and the head of the caudate nucleus. In **study 3** we repeat the behaviour performance, introducing a preference classification instruction to further the cognitive system into an implicit learning mode, and report a clear and increasing preference for grammatical structure over repeated sessions. In **study 4** we investigated the basal ganglia component in Huntington patients with specific caudate head lesions. While the patients did not show any deficit in their behaviour

performance, structures in the basal ganglia including the caudate head showed abnormal activation patterns compared to their matched normal controls. Also, a cooperative activation between basal ganglia and hippocampus typically involved in declarative memory was found. We interpret this to reflect attempts of the cognitive system to compensate the damaged procedural processing with declarative knowledge processing. In summary, in the studies of this thesis we have gained an initial characterization of the neural infrastructure subserving artificial grammar processing. We have done so by characterising the end-state of the learning process as well as characterizing the learning curves reflecting the outcome of acquisition at different time points. This thesis reports findings supporting the view that the extended artificial grammar learning model is useful to capture structural aspects in language acquisition processing in the laboratory.

LIST OF PUBLICATIONS

- I. Petersson, Karl Magnus, Forkstam, Christian, & Ingvar, Martin (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, 28, 383-407.
- II. Forkstam, Christian, Hagoort, Peter, Fernandez, Guillen, Ingvar, Martin, & Petersson, Karl Magnus (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage*, 32, 956-967.
- III. Forkstam, Christian, Elwér, Åsa, Ingvar, Martin, & Petersson, Karl Magnus (2008). Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research*, 1221, 80-92.
- IV. Forkstam, Christian, Voermans, Nicol, Dekkers, Marieke, Kremer, Berry, Fernández, Guillen, & Petersson, Karl Magnus (In submission). Frontostriatal circuitry in artificial grammar learning: An FMRI study in Huntington's disease.

CONTENTS

CHAPTER 0 — PREFACE.....	1
CHAPTER 1 — BACKGROUND.....	4
1.1 Objectives	4
1.2 Artificial Grammar Learning	6
1.2.1 Outline of the artificial grammar learning paradigm	6
1.2.2 Background	9
1.2.3 Implicit sequence learning and dual mechanisms.....	11
1.3 Knowledge acquisition	14
1.3.1 Objective 1	14
1.3.2 Adaptive cognitive systems	14
1.3.3 Learning as a cumulative function.....	17
1.3.4 The perception-action cycle.....	17
1.3.5 Learning paradigms.....	21
1.3.6 Cognitive models	22
1.3.7 Levels of description of cognitive processes.....	26
1.3.8 Memory systems	27
1.3.9 Implicit sequence learning.....	30
1.3.10 Knowledge representation and learnability	33
1.4 The language faculty	39
1.4.1 Objective 2	39
1.4.2 Language acquisition	39
1.4.3 The language acquisition device.....	41
1.4.4 Formal grammars	43
1.4.5 The Reber grammar	45
1.4.6 Acquisition of formal grammars and artificial grammar learning	48
1.5 Neuronal implementation.....	50
1.5.1 Objective 3	50
1.5.2 Structural and functional complexity of the brain.....	50
1.5.3 Frontal lobe function.....	52
1.5.4 The left inferior frontal region.....	54
1.5.5 Functional anatomy of the corticostriatal circuits	57
1.5.6 Neural correlates of sequence processing	60
1.6 Summary.....	65
CHAPTER 2 — METHODOLOGICAL BACKGROUND	68
2.1 Whole brain investigations.....	68
2.2 Methods	69
2.3 The coupling between neural activity and regional cerebral blood flow	70
2.4 Image processing and statistical analysis.....	71
2.4.1 Image pre-processing.....	73
2.5 Statistical modeling and estimation	75
2.5.1 The general linear model	75
2.5.2 Baseline fluctuations and global normalisation	76
2.5.3 Choosing reference state	77
2.6 Hypothesis testing and statistical inference.....	78
2.7 Voxel based morphometry	80

CHAPTER 3 — Specific background to the included studies.....	82
3.1 Stimulus material.....	82
3.2 Experimental procedure.....	83
3.3 Behaviour data analysis.....	83
CHAPTER 4 — Overview of the experiments	85
4.1 Study 1: Artificial syntactic violations activate Broca’s region.....	85
4.2 Study 2: Neural correlates of artificial syntactic structure classification	87
4.2.1 ERP correlates of artificial syntactic structure classification	93
4.3 Study 3: Instruction effects in implicit artificial grammar learning: A preference for grammaticality.....	98
4.4 Study 4: Frontostriatal circuitry in artificial grammar learning: An fMRI study in Huntington’s disease	102
CHAPTER 5 — Conclusions.....	106
CHAPTER 6 — Acknowledgements.....	109
CHAPTER 7 — References.....	110
CHAPTER 8 — Included studies	132

LIST OF ABBREVIATIONS

ACS	Associative chunk strength
AGL	Artificial grammar learning
BA	Brodman's area
BOLD	Blood-oxygen-level dependent
EEG	Electroencephalogram
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
HD	Huntington's disease
MR	Magnetic resonance
MRI	Magnetic resonance imaging
G/NG item	Grammatical/Non-grammatical item
H/L item	Item with high/low ACS
HG/HNG item	Grammatical/Non-grammatical item with high ACS
LG/LNG item	Grammatical/Non-grammatical item with low ACS

CHAPTER 0 — PREFACE

An expert is a man who has made all the mistakes which can be made, in a narrow field.

Niels Bohr

The definition of insanity is continuing to do the same thing and expecting a different result.

John Arquilla, on the way America fights terrorism

ehm - well, alltfoer maanga idioter under alla omstaendigheter - aeven om det staat taemligen klart foer oss andra att vi inte foeds som oskrivna blad aer detta tydligen ingen garanti foer att inte bli ett.

Anonymous, 2007

Vad menar vi när vi säger att vi tänker? För att kunna uttrycka eller skriva ner en fråga som denna utnyttjar vi vårt språks möjlighet att formulera och kommunicera ett meningsbärande budskap. Vi tar till oss budskapet när vi hör eller läser frågan, förutsatt att vi delar samma språk. Vi kan skapa mentala bilder för att visualisera för oss själva hur vi upplever det när vi tänker en tanke. Vi förstår intuitivt, eller kanske bara tror oss förstå, hur känslan av att tänka upplevs för oss. Vi kopplar upp oss till diverse analysverktyg som medan vi utför den mentala processen att tänka mäter effekten av hjärnans informationsprocesser i form av t ex syrekonsumention i olika delar av hjärnan, eller förändringar över tid i elektricitetsfältet som hjärnan genererar utanför skallbenet. Dock ger varken den subjektiva upplevelsen vi bär när vi introspekterar vår tanke eller den objektivt mätbara analysen av densamma ett tillfredställande svar på frågan. Kanske är ett fullständigt tillfredställande svar orimligt. Kanske är det enda som kvarstår vår förmåga att uttrycka frågan.

Strukturer för tänkande

Genom att analysera en funktion som understödjer flera mentala kapaciteter kan en underliggande gemensam neuronal arkitektur studeras. Inläring av konstgjord grammatik (så kallad *artificial grammar learning*) har kraft att nå detta mål. Studiet av artificiell grammatikinläring brygger mellan mentala processer och funktioner, såsom olika minnesfunktioner (procedurell inläring, icke-motorisk och implicit sekvensinläring, etc.)

och språkfunktioner (t ex (morpho-) syntaktiskt processande). Förståelsen av uppbyggnaden och mekanismerna som understödjer inläring och processande är viktig för en förståelse av mänsklig kognition.

Procedurell inläring är beteckningen för förmågan att sammankoppla en sekvens av stimulus-responspar för att lära sig att observera eller genomföra en sekvens eller mönster av händelser. Denna typ av inläring sker utan att individen kan uttrycka förmågan verbalt, åtminstone i ett initialt skede av inlärningsprocessen. Sluttillståndet kallas procedurell kunskap, alltså kunskap om hur man genomför en viss färdighet. Detta ska särskiljas från deklarativ kunskap som är förmågan att avgöra om någon proposition är korrekt. Procedurell kunskap sträcker sej utöver enkla motoriska färdigheter. Man kan betrakta många av våra mentala förmågor som olika exempel av procedurella färdigheter.

En del av språkfunktionen utgörs av de regler eller strukturer vars uppgift är att avgränsa de fonemsträngar som hör till språket, dvs syntax. Antagligen kommer mer än en uppsättning av grammatiska konstruktioner och vokabulär att kunna generera samma totala uppsättning strängar (Quine, 1992). Det är denna funktion som bryter ner mångdimensionella tankematriser till ett endimensionellt seriellt ordflöde, för att åter bygga upp dem hos mottagaren. Oavsett om den syntaktiska protofunktionen är arts specifikt medfödd eller ej, så utvecklas den genom inläring under individens uppväxt och liv. Var och en av oss lär sig sitt språk genom att observera andra människors verbala beteenden, och genom att få sitt eget bristfälliga verbala beteende observerat och förstärkt eller korrigerat av andra. Vi är helt beroende av tillgängligt beteende i observerbara situationer. (Quine, 1992). Denna avhandlings studier har ett värde för vår förståelse av denna syntaktiska funktions neuronala realiserande.

As a general disclaimer, there is for me no claim of originality in writing this thesis. All possible misrepresentations of concepts, or for that matter, failure of understanding the ideas of others, are unintended. What further is, the concepts and ideas included in this thesis are not new but are well-known in either the field of cognitive neuroscience or in its related disciplines. I base much of my knowledge on the matter from the excellent works of those researcher that followed similar tracks in their investigations, as well as the many discussions and helpful guidance from those researcher that has surrounded me, and, intensionally or not, acted as my supervisors. Specifically, this work would not have been pursued without the kind supervision from my two benefactors, Martin Ingvar and Karl Magnus Petersson, who

made it possible for me to work and study not only at the PET and MR facilities at the Karolinska University Hospital, but also in the cognitive laboratories at the Donders Institute and the Max Planck Institute for Psycholinguistics in the Netherlands.

This thesis is based on an empirical quest, centred on the use of the indirect measure of brain activity using fMRI, to motivate the use of a simplified paradigm for implicit learning as a model for language acquisition and/or language processing. At the time of the start of this quest it was to my knowledge the first attempts to investigate artificial grammar learning in the normal healthy young adult brains using functional neuroimaging. However, and as expected, as we looked further into the task layed out for us we found others to be were with us on the journey. The empirical works on which the thesis is based includes cognitive-behavioral laboratory experiments in combination with functional neuroimaging methods. The thesis is composed of seven chapters, a reference list, and a last section with the four papers which are the basis of the thesis. The thesis starts out with an introduction to the artificial grammar learning paradigm followed by a general background (chapter 1), and methodological background (chapter 2 and 3) for the experimental studies that are discussed in chapter 4. The first experimental study discussed in chapter 4 represent our first attempt to investigate the brain regions involved in implicit sequence learning. The second study confirms and expands beyond the findings in the first study. The third study is constructed as a sequence of behaviour investigation comparing the classical artificial grammar learning paradigm and a version working on mere exposure using yes/no classifications. This modification is then used in study 4 in an investigation on Huntington's disease patients and healthy controls, in which we tried to clarify the involvement of the caudate nucleus of the basal ganglia during classifications found in study 2.

Investigations on different memory systems have mainly concentrated on different aspects of declarative types of memory but not that much on the unconscious variants such as procedural memory and mental skill acquisition, and more effort has been made to study procedural motor learning than non-motor learning (Seger, 1998). The focus of this thesis, on non-motoric procedural learning, is an example of investigations correcting this imbalance.

Christian Forkstam

2009-12-24

CHAPTER 1 — BACKGROUND

*– Är ett oskarpt fotografi överhuvudtaget någon bild av en människa?
Ja, är det alltid en fördel att ersätta en oskarp bild med en skarp? Är det inte
ofta just det oskarpa som vi har behov av? Man kan t ex säga “Uppehåll dig
ungefär här”.*

(Wittgenstein, 1953, Filosofiska undersökningar, p. 45)

1.1 OBJECTIVES

The general objectives with the studies of this thesis were to develop an information processing model for the cognitive functions supporting implicit sequence learning. The thesis presents empirical and theoretical background on non-motor implicit sequence learning. It provides an introduction to recent developments on artificial grammar learning, a standard investigative paradigm on implicit sequence learning. In specific, the thesis introduce the implicit artificial grammar learning paradigm and evaluate its use for investigating implicit sequence learning in general, and language acquisition and processing in specific.

The nature of the knowledge acquired during the implicit learning of an artificial grammar, its knowledge representation, and the functional role this knowledge acquires during learning, is useful to understand from a cognitive neuroscience perspective in general. Such pursuit includes the characterization of the neural infrastructure subserving these aspects of artificial syntax processing, both the end-state of the learning process as well as the learning process itself, using different functional neuroimaging approaches and behaviour measures. This will make it possible to compare natural and artificial language processing in the human brain, as well as to address questions related to the characteristics of the learning mechanism(s) involved, the nature of the knowledge acquired, how this knowledge is represented and how it is put to use.

The objectives with this thesis are to investigate the functional mechanisms of implicit acquisition and classification of rule based implicit sequence processing. We will in this thesis investigate artificial grammar learning in order to:

1. Explore the underlying neural processing architecture for implicit learning/acquisition to acquire and process non-motor sequences, an implicit non-motor procedural learning ability present in the human cognitive system (**section 1.3**),
2. Validate and explore the repeated artificial grammar learning paradigm as a laboratory model to investigate the acquisition and processing of structural aspects of language, e.g. (morpho-) syntax processing (**section 1.4**), and
3. Further our understanding of the specific neural processing architecture which subserves the language faculty and its syntax processing ability (**section 1.5**).

1.2 ARTIFICIAL GRAMMAR LEARNING

All studies included in this thesis make use of artificial grammar learning to investigate implicit sequence learning processes in the human. They all follow the classical artificial grammar learning paradigm as outlined in Reber (1967), the seminal study on artificial grammar learning. In this study participants memorized sets of letter strings such as TPPTS or V XVPS. Unbeknownst to the participants, the strings were generated by traversing through a finite-state grammar similar to the graph in **Figure 1.1**. After the study phase participants were informed that the strings they had just memorized conformed to a complex set of grammatical rules. They then received new strings that were either grammatical or not, and were asked to make a grammaticality judgment for each string based on the strings they had studied earlier. An average of 69% of the classification decisions was found to be correct. He proposed that classification was based on abstract rules that participants had formed implicitly during the initial study phase, that is, without using conscious, verbalisable strategies. The acquired rules were regarded as tacit knowledge, that is knowledge unavailable for conscious inspection (Reber, 1989).

We will in this first chapter introduce the artificial grammar learning paradigm, followed by background on knowledge acquisition and implicit sequence learning (**section 1.3**), the language faculty and syntax processing and acquisition (**section 1.4**), and neuronal implementation of the processes that support implicit sequence learning (**section 1.5**).

1.2.1 Outline of the artificial grammar learning paradigm

The artificial grammar learning paradigm used throughout the thesis is a paradigm divided into an extended acquisition phase and several classification phases. The purpose with the acquisition phase is to covertly expose the subject to the underlying regularities of a specific grammar (**Figure 1.1**), and the purpose of the classification tests is to quantify the subjects' level of implicit learning over the acquisition period.

During the acquisition phase the participants are occupied with short-term memory tasks where they keep in active memory a set of letter strings that follow the sequential construction of a grammar. The acquisition task is set up as several short-term memory task distributed over the acquisition days. It consists of immediate serial recall of each letter string by typing the string correctly on a keyboard, completely without performance feedback. The acquisition set is composed of grammatical strings only, sometimes referred to as positive examples of the grammar.

In one type of classification the subjects are informed that the acquisition strings were generated by a complex set of rules. They are then asked to classify previously not encountered items based on their immediate intuitive impression (i.e., to guess based on ‘gut feeling’), as violating these rules or not. This type of instruction is called *grammaticality instruction* and the test *grammaticality classification*.

In some of the studies manipulation of the instruction has been used with the aim to minimize the potential that the subjects develop problem solving strategies during classification, potentially enforcing explicit learning processes in parallel with ongoing implicit learning processes. In this alternative type of classification the subjects are instead asked to classify new items, not previously encountered, according to whether they like them or not. Here, just as in the grammaticality classification, they are instructed to classify based on their immediate intuitive impression (i.e., to guess based on ‘gut feeling’) and they are informed that this strategy would yield the best performance. This type of instruction is called *preference instruction* and the test *preference classification*. Preference classification take advantage of the structural mere exposure effect (Manza & Bornstein, 1995). In mere exposure artificial grammar learning subjects receive a preference classification instruction which make no reference to any previous acquisition episode and the subjects are not informed about the existence of an underlying generative mechanism. The idea is that mere exposure artificial grammar learning might measure implicit knowledge in a more pure manner because there is nothing in the classification procedures that refers to the acquisition part of the experiment and no reference to a complex set of rules are ever made. In both preference and grammaticality classification, performance lies reliably above chance (Forkstam, Elwér, Ingvar, & Petersson, 2008; Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Petersson, Forkstam, & Ingvar, 2004; Reber, 1967; Stadler & Frensch, 1998). At the same time verbal reports reveal no or little explicit knowledge, indicating that implicit learning of relevant information related to the grammar has taken place.

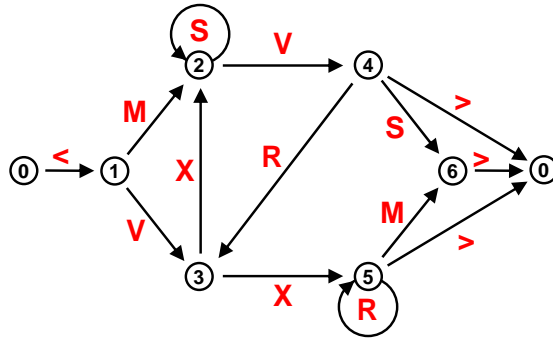


Figure 1.1 Transition graph representation of the Reber machine (similar to a program machine table) used as generative system to construct grammatical (e.g., <MSSVRXR>) in contrasts to non-grammatical strings (e.g., <MSMVRSR>) used throughout this thesis.

The classification stimulus material is composed in a factorial design with the factors grammaticality and associative chunk strength (ACS). Associative chunk strength is a statistical measure of substring familiarity between a given classification item and the acquisition set of items, in specific, the similarity of the classification items to the acquisition set in terms of 2 and 3 letter substring frequencies (see **Table 1.1**). During classification sensitivity to ACS offers a mean to quantify sensitivity to the frequency-count of chunks of letters in relation to previously encountered acquisition strings. In this way the artificial grammar learning paradigm can empirically quantify both the level of implicit acquisition based on grammatical examples, as well as independently quantify the level that familiarity to local substrings influence the classification performance. This design makes it possible to independently investigate processes working on structural knowledge and processes based on local substring familiarity.

Associative chunk strength (ACS)		
2x2 factorial design	High ACS	Low ACS
	25% HG	25% LG
Grammatical	(High ACS/Grammatical)	(Low ACS/Grammatical)
Non-grammatical	25% HNG	25% LNG

Table 1.1. Structural knowledge vs. Local substrng familiarity. Outline of the factorial design composition of the classification stimuli material.

1.2.2 Background

To understand the human capacity to learn to communicate through language is an outstanding scientific challenge (Chomsky, 2000; Hauser, Chomsky, & Fitch, 2002; Jackendoff, 2002). Chomsky has frequently quoted von Humboldt's description of language as a system which *infinite use of finite means*, meaning that an infinite number of sentences can be created using a finite number of grammatical rules (e.g., cited in Chomsky, 1965). Since the 1950s a fundamental problem in theoretical linguistics has been to construct explicit models reflecting this intuition (Chomsky, 1965; Newmeyer, 1995). The simplest formal model incorporating the idea of infinite use of finite means is represented by the family of regular grammars: right-linear phrase structure grammars (Chomsky, 1957). The Reber grammar used throughout the investigations included in this thesis is an instantiation of such a regular grammar (see **Figure 1.1**).

Investigations on implicit sequence learning suggests that humans are equipped with mechanisms for learning or acquisition with the capacity to extract structural information implicitly from the experience of observed exemplars, without induction of an explicit model (Reber, 1967; Stadler & Frensch, 1998). It has been suggested that such acquisition mechanisms play an important role in several types of information extraction processes or forms of learning (e.g., Cleeremans & McClelland, 1991; French & Cleeremans, 2002; Lewicki, 1986; Stadler & Frensch, 1998). The task of learning an artificial grammar has previously been used as a laboratory model for investigating aspects of language learning in infants (Gomez & Gerken, 2000), exploring key differences between human and animal learning relevant to the narrow faculty of language (Hauser et al., 2002), as well as second language learning in adults (Friederici, Steinhauer, & Pfeifer, 2002). That humans can learn artificial grammars in an implicit fashion was shown in the seminal study of Reber (1967), who suggested that relevant information of the rules of the grammar was abstracted from the environmental input. He suggested further that this process represented a mechanism that is intrinsic to natural language learning.

Several theories on artificial grammar learning has been put forward since before the study of Reber (1967), in which the acquired knowledge was suggested to be best represented as a set of rules. In the theory of rule-based cognitive representations, participants are thought to, implicitly, acquire cognitive representations of the rules of the grammar used to generate the stimuli, and use these rules for grammaticality judgments. In an opposing theory, the exemplar-based account, participants are thought to store exemplars during the learning phase, and then base their grammaticality judgments on some relevant whole item similarity measure. In the fragment or chunk (n-gram) account, it is thought that distributional properties of local sequential regularities/contingencies are acquired during learning, and that subjects base their judgment on measures of associative strength of the chunks present in a given string, in relation to the frequency of these chunks collapsed over all previously experienced strings. In the dual mechanism account, a combination of these latter theories was joined: simultaneous acquisition of both abstract rules and fragment learning. Finally, in the episodic processing account, it is thought that episodic processing knowledge is acquired during learning in addition to structural knowledge (e.g. rules or fragment statistics), which can be both explicit or implicit. Here, the training instructions and the learning task determine which aspects of the training items are encoded, while classification performance is determined by the degree to which test instructions reinstate the processing context during acquisition (cf. Johnstone & Shanks, 2001; Shanks, 1994).

The computational architectures suggested to subserve artificial grammar learning includes both classical and connectionist types. In the production-rule classifier architecture, conditions are suggested to code for sequence features encoded probabilistically during learning, and that actions interact competitively during classification. In the exemplar-similarity architectures, it is thought that exemplars are encoded during learning and that classification is based on neighbourhood similarity measures. In architectures based on hierarchical competitive chunking, it is thought that elementary chunks are hierarchically joined into non-overlapping higher-order chunks during learning. Classification is then based on the number of chunks necessary to represent a given sequence, that is, a measure of familiarity. In the feed-forward neural network architectures, an input-output mapping is learned in an unsupervised fashion, that is, it encodes co-occurrence or sequential contingencies between input features. Classification is based on a goodness-of-fit measure between input and output during testing. Finally, in recurrent neural network architectures, for example the simple recurrent networks, distributional properties of sequence fragments (n-grams) are acquired in an unsupervised fashion. Classification is here based on the overall prediction performance during testing.

1.2.3 Implicit sequence learning and dual mechanisms

We will now shortly summarize some from my opinion relevant studies which all have in common that they support the idea of dual mechanism at play in artificial grammar learning (Forkstam & Petersson, 2005). In a study by Knowlton and Squire (1996), amnesic patients and normal controls were investigated on both a classical and a transfer version of the artificial grammar learning task, in which the subjects classified grammatical and non-grammatical items composed from a novel letter alphabet. The normal controls and the amnesic patients performed similarly on both artificial grammar learning tasks and the amnesic patients showed no explicit recollection of whole-item or fragment information. Knowlton and Squire suggested that these results indicate that the explicit recollection in the normal controls reflects an epiphenomenon not necessary for adequate performance on grammaticality classification. Instead, artificial grammar learning depends on the implicit acquisition of both abstract (i.e. rule-based) and exemplar-specific information related to associative chunk strength (ACS).

In a study by Meulemans and Van der Linden (1997), the initial acquisition phase, using a relatively small set of acquisition items, was characterized by the acquisition of distributional local sequential (fragment/chunk) information in the training set, since the initial performance correlated with ACS. When instead using a relatively large acquisition set, they suggested that the late acquisition phase is characterized by the abstraction of grammatical structure, since the performance did not correlate with ACS. Additionally, in neither phase did the performance correlate with explicit recollection as measured in a production test.

Both these studies suggest that artificial grammar learning is dependent on implicit learning mechanisms as originally proposed by Reber (1967). However, while Reber (1967) originally argued that the implicit learning process abstracted rule-based knowledge, these more recent studies indicate that dual mechanisms might be at play (for an alternative perspective see Johnstone & Shanks, 2001). It is clear from these studies that distributional information of local sequential regularities are acquired and used for grammaticality classification (at least during certain stages in the learning process). Knowlton and Squire (1996) argued for abstract representations (i.e. rule-based representations) based on the performance of the amnesic patients and their normal controls on the transfer version of the artificial grammar learning task. However, it is not clear that this conclusion follows from the results on the transfer task. Specifically, this conclusion appears to depend on the assumption that transfer performance is

critically dependent on abstract representations. This is not necessarily the case. Transfer performance is dependent on a mapping from the representation of the acquired knowledge to the new surface form, which by necessity has to be established during the initial phase of the transfer task, and it is not clear that it is easier to generate such a mapping from an abstract knowledge representation in comparison to a surface based knowledge representation. On the other hand, the results of Knowlton and Squire (1996) is compatible with an abstract representation and the observation that the classification performance did not correlated with ACS when the participants had reach the late acquisition phase indicate that abstraction of grammatical structure in do take place.

In a study by Chang and Knowlton (2004) it was shown that the sensitivity to grammaticality status was not affected by a change in low-level visual features (font/case), while this change reduced the sensitivity to ACS. The addition of a secondary task (articulatory suppression) during the acquisition phase eliminated font sensitivity and reduced the contribution of ACS to the classification performance. Consistent with FMRI data, showing reduced activation levels for high vs. low ACS items in early visual regions (Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004), some aspects of classification performance might be related to perceptual fluency (e.g., repetition priming), since changes in surface features reduced the ACS sensitivity. In an event-related FMRI study, using a balanced chunk strength design, Lieberman and colleagues (2004) reported several brain regions that may contribute to artificial grammar learning classification. The main findings suggested that the caudate nucleus was more active for grammatical vs. non-grammatical items, while the medial temporal lobe (MTL) seemed to be more active for high vs. low ACS. Moreover, they reported some evidence for a negative correlation between the caudate and hippocampal activations, which they interpreted as suggesting a competitive relationship between the two regions. However, the observation of a negative correlation does not necessarily imply a competitive relationship and there is also some evidence suggesting that the MTL and the caudate nucleus can interact cooperatively (Voermans, Petersson, Daudey, Weber, van Spaendonck, Kremer, & Fernández, 2004).

In a series of artificial grammar learning experiments, Zizak and Reber (2004) examined the links between the classic- and structural mere exposure effects (i.e., the preference of previously encountered to novel items and the acquisition of syntactic/structural regularities by being exposed to stimuli, resulting in a positive correlation between preference and grammaticality status on new items, respectively). In these experiments, subjects either classified stimuli based on grammaticality or rated them in terms of likeability. The grammar

was instantiated with familiar and unfamiliar symbols, and participants showed standard artificial grammar learning effects in all cases. However, whether the two exposure effects were dependent on symbol familiarity (high familiarity produced the structural mere exposure effect; moderate familiarity produced only the classic mere exposure effect; unfamiliar symbols produced neither exposure effect). In another series of behaviour experiments, Domangue and colleagues (2004) argued that learners can use at least two types of knowledge: an explicit model or instance memories. The subject performance was characterized in terms of response times and accuracy with respect to their ability to generate letter sequences. The acquisition conditions were experimentally controlled in order to manipulate the availability of the two types of knowledge. The memory based condition yielded rapid response times but less accuracy compared to the model-based acquisition condition.

In a study of implicit acquisition of acoustic regularities (transition probabilities between timbres), Tillmann and McAdams (2004) extend previous implicit learning results to the domain of complex nonverbal auditory material. Their results suggest that listeners become sensitive to statistical regularities independent of acoustical surface characteristics in the stimulus material. A study by Bitan and Karni (2003) showed that readers are able to acquire an artificial script both with explicit and implicit acquisition instructions. In a follow-up study, Bitan and Karni (2004) provided data suggesting that letter decoding can evolve from implicit training on whole-word recognition and that the acquired knowledge was independent of explicit letter knowledge (measured by declarative recognition). They concluded that both implicit (procedural) and explicit (declarative) knowledge contributed to letter decoding and word-specific recognition, suggesting the dependency on explicit knowledge as related to the possibility that both routines become proceduralised with practice.

1.3 KNOWLEDGE ACQUISITION

Human learning abilities are to a large extent unconscious or implicit in nature, similar to other primates and many other mammals. Many of these constitute evolutionary old features still shared with other vertebrates. Even though there is great variability between different implicit learning abilities they all share certain features such as fast computation and rigidity, that is, they are not easily modified. The combination of these features is of fundamental importance for a slowly adaptable system with robust characteristics to endure external stress on the system. Conscious learning abilities are on the other hand comparably fast, are less rigid and can be much more easily modified than unconscious ones. These properties provide a quickly adaptable system at the expense of stability. Since the beginning of conscious learning abilities, both conscious and unconscious learning systems have co-existed in the brains that evolved in the human primate lineage.

Learning is defined in this thesis as adaptive processes by which the brain functionally restructures its processing pathways/networks or its representations of information as a function of experience. The memory trace (i.e., the stored information) is the resulting changes in the processing system due to learning. On this view, learning is a dynamic consequence of information processing and network plasticity. From this perspective, and in contrast to simple information storage, learning and adaptation can be viewed as a process of generalisation.

1.3.1 Objective 1

We will through investigation on artificial grammar learning explore the underlying cognitive processing architecture for implicit learning/acquisition to acquire and process non-motor sequences, an implicit non-motor procedural learning ability present in the human cognitive system.

1.3.2 Adaptive cognitive systems

A cognitive system such as the human brain in interaction with its current environment is—like any other biological system—the outcome of its evolutionary and developmental history. Both its individual history (ontogenesis) and its evolution past (phylogenesis) has to be taken into account to understand its different characteristics. The capacity of an embodied cognitive system to develop and learn provides a necessary basis for cultural and evolutionary interaction. From the view of information processing systems, cognition is equated with internal information processing. A cognitive system is interfacing with the external

environment and its processing sub-system is interfacing with other sub-systems, internal components that receives input from and transmits output to other sub-systems.

A physical system is an information processing device when its physical states can be viewed as cognitive-representational (in the sense of Jackendoff, 2002) and transitions between these can be conceptualized as a process operating on these cognitive structures. Information processing, or state transitions, can be conceptualized as trajectories in a state space. The brain can only represent information in the terms of numbers, that is, in terms of membrane potentials, inter-spike-intervals, or any appropriate set of dynamical variables, and therefore cognitive structures are not represented in the brain in a simple transparent manner. It is possible to simulate all finitely specified symbolic cognitive models as processes on numbers. These models can all be emulated in dynamical systems such as recurrent neural networks (for a review see, Siegelmann, 1999).

From a parallel distributed processing perspective, learning in a neural network is a dynamic consequence of information processing and network plasticity. Brain regions do not merely act locally but interact with one another in complex neural networks (Churchland & Sejnowski, 1993). Learning is in this manner depends on changes in the interactions between regions in a neural network.

Memory is a process decomposed into several processing stages, including on-line encoding (i.e., representation of the information to be stored), memory formation and storage, consolidation, re-organisation and maintenance, as well as retrieval. Different acquisition problems require different learning processes, instantiated in various memory systems, in order to ensure efficient solutions to learning problems. These processing systems, with multiple interacting memory systems, operate at several different time-scales, spanning some seven to nine orders of magnitude. An independent rationale for the existence of multiple memory systems is related to the serial learning problem, also called the stability-plasticity dilemma, i.e. the problem to update the knowledge base by storing and integrating novel information with previously acquired information. There is a trade-off between stability and plasticity. Stability is necessary to ensure robust process reliability, while plasticity is necessary for the acquisition of new information. Too much stability precludes sufficient plasticity, and conversely, too much plasticity threatens processing stability.

Learning in the formulation of information processing systems with adaptive properties is conceptualized as the interaction between two or more sets of dynamical variables, cognitive-representational and adaptive, respectively. Another form of memory can be instantiated in the state-space by the representational dynamics alone. This is a form of process memory which does not depend on learning instantiated in adaptive parameters. Fundamentally this form of memory is related to the fact that the current state on the state-space trajectory can be seen as representing aspects of the systems processing history. This form of process memory depends on the topology of possible state-space trajectories. This is the form of memory that finite-state architectures can instantiate and is commonly used as a strategy to implement short-term memory properties in such processing systems (i.e., via state-space coding, cf. e.g., Hopcroft, Motwani, & Ullman, 2000). It follows that the processing system's response to a given input depends on the current internal state, where the current internal state may also represent the process history.

Several memory researchers have argued on both theoretical and empirical grounds that the brain is equipped with multiple memory systems (e.g., Eichenbaum & Cohen, 2001; Schacter & Tulving, 1994b; Squire, Knowlton, & Musen, 1993; Stadler & Frensch, 1998). These memory systems serve different purposes and are therefore thought to store different types of information. Tulving (1995) suggested that cognitive memory research can be meaningfully divided in the two general concepts of memory systems and memory processes. Tulving proposed a simple model for memory organisation, in which cognitive memory systems are related to one another in terms of memory processes supporting encoding, storage, and retrieval, the SPI model. The SPI model is built on the hypothesis that all memory systems are related in a process specific manner in the sense that information is serially (S) encoded into the systems, memory traces are stored in parallel (P), and information can be retrieved independently (I) from each memory system. An important claim of the SPI model is that the relations between different memory systems are process specific. The serial character of encoding is consistent with the principle of adaptive changes and information processing in the brain. Information encoded simultaneously in several interacting memory systems does not imply that these interactions are best conceptualized as serial. Instead, there are dependencies between different processing systems, and in this sense the encoding of information in one system might be contingent on the processing of information in some other memory system. The output of one system is input to another.

1.3.3 Learning as a cumulative function

Learning is cumulative in the sense that the effects of experience are carried forward to improve later performance, a property fundamental for the construction of learning curves. To justify the use of learning curves we plot the sampled data points in a xy co-ordinate space where the x -axis represents the independent variable and the y -axis the dependent response measure. This determines the discrete data points to be connected. We then find a rule for operationally defining a function, $y = f(x)$, that describes a smooth curve that joints overall the data points. This is the function whose existence needs to be justified. Without it only discontinuous plots (e.g., bar graphs) would be justified. When designing a learning experiment, we can assume a vague, intuitive rule presumed to connect the trials into a series even if no such rule exists for the learner to follow, or if a different rule is actually used by the learner. Because of this possibility we might assume a simpler basis to the continuity over successive trials than is actually the case. It therefore cannot be enough simply to operationally define a function by which a curve might be drawn to fit observed data points, but one must also show why that particular function is a reasonable candidate for modelling the rule-governed constraints actually responsible for the learning accomplished by the organism in a given context.

1.3.4 The perception-action cycle

Perception is roughly a mapping from the series of values taken by the environmental variable into the series of values taken by the organism variable, whereas action consists of the inverse mapping. The individual brain interfaces with its environment through sensory input and motor output, in a perception-(cognition-)action cycle: sensory input \rightarrow perceptual processing \rightarrow (cognitive processing \rightarrow) temporal organisation of motor output \rightarrow action (Equation 1.1, Figure 1.2), represented here as:

$$P(t_n) \rightarrow A(t_{n+1}) \rightarrow P(t_{n+2}) \rightarrow \dots \quad (1.1)$$

where $n = 0, 1, \dots$, P = perception, A = action. The brain receives perceptual information through several sensory modalities and coordinates actions in the form of movements of the skeletomuscular apparatus, the autonomic nervous system, and other output systems such as the larynx and tongue. The perception-cognition-action cycle emphasises the overarching heuristic function of the brain (Rosenbleuth, Wiener, & Bigelow, 1943; Wiener, 1948), which both supports an individual survival and success within a particular ecosystem, increasing the likelihood of the genetic information of this individual to be passed on to the genetic pool of

the species. Perception and action operate on each other reciprocally, in a fashion that can be formulated as a spiral through space and time. This co-variation that results involves both environmental and organismic variables.

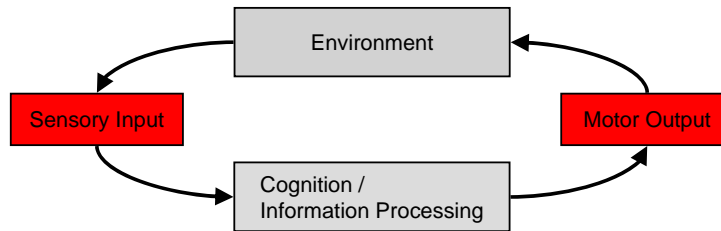


Figure 1.2 The perception-cognition-action cycle. The perceptual systems allow the human brain to extract relevant patterns of information from, at times, a noisy, changing, and unpredictable environment, while the motor output apparatus allows it to temporally organise behaviourally relevant actions and act in a goal directed fashion in its environment (including e.g. the creation of artefacts, communicating with conspecifics, as well as to effect changes in the physical and socio-cultural environment). Here cognition and internal information processing is equated.

Brain complexity is reflected in the structural composition of its processing units—neurons. This includes the dendritic tree and neuronal soma composition, axonal arborisation, synaptic organisation and passive as well as active membrane properties supported by voltage- and neurotransmitter-gated ion-channels. These characteristics provide neurons with adaptable nonlinear dynamical properties (Koch, 1999; Shepherd, 1997). Chemical synapses show a number of different forms of plasticity with characteristic time-scales that span at least nine orders of magnitude, from milliseconds to weeks, providing a necessary substrate for learning and memory (Anderson, 2002; Koch, 1999; Koch & Laurent, 1999).

Information is received through the input synapses of a neuron and flows from the dendritic tree, via the soma, to the axon hillock affecting the local membrane potential positively or negatively. Once triggered, the action potential spread along the axon and the final terminal arborisation, where neurotransmitters are stochastically released into the synapse from the pre-synaptic membrane, affecting the synapse environment and the post-synaptic receptor

dynamics. The increased concentration of neurotransmitters causes them to diffuse across the synaptic cleft and activates post-synaptic receptors generating a post-synaptic potential, which starts the whole process anew in the downstream neuron. In all its roles, the nervous system invokes neuronal processing, to store information through memory formation and changes in its adaptable properties given its dynamic processing environment. The perception-cognition-action cycle therefore needs to be complemented by the encoding-storage-retrieval cycle (**Figure 1.3**).

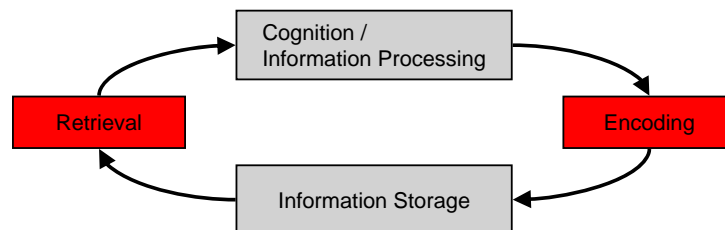


Figure 1.3 The encoding-storage-retrieval cycle. Learning can be defined as the processes by which the brain functionally restructures its processing networks and/or its cognitive representations as a function of experience. The stored information is the resulting changes in the processing system. The processing system is non-stationary and learning in a neural network is the dynamic consequence of processing and network plasticity.

The perception-action cycle and the encoding-retrieval cycle interact through active processing of information subserved by various forms of short-term working memories. There is not only one encoding-retrieval cycle but several, and likewise, that there are several parallel perception-action cycles. This gives rise to the idea of different memory systems as well as interacting cognitive modules (**Figure 1.4**).

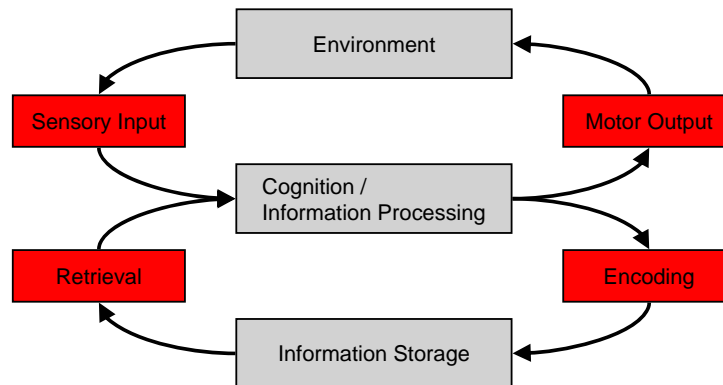


Figure 1.4 Interaction between the perception-cognition-action and the encoding-storage-retrieval cycle. To incorporate the capacities for memory, learning and adaptation, the perception-cognition-action cycle is complemented with the encoding-retrieval cycle interacting through active information processing in for example working memory, such that adaptation and learning is a functional consequence of information processing.

By endorsing the principle of organism-environment interaction one has to consider two sets of the data points, one showing how the learner's perception of the task variables changed over time and one showing how the learner's action (response characteristics) changed as a function of the task variables. Similarly, two learning curves must be justified, one for perceptual learning and the other for action learning. Two mathematical rules must be found for determining how the data points are to be connected to form a series, and two laws must be found to justify use of the pair of functions.

Generally, we define learning as a cumulative function L that determines a mapping between two series: a *perceptual* (stimulus) series and an *action* (response) series. The perceptual series consists of episodes (e.g., trials) on which the learning function applies to increase the value of some response characteristic of the organism over time.

Learning is not merely a simple increase of a response tendency with repeated experience. Often, if not always, specific and non-specific changes in the general disposition to learn also accompany these changes in response characteristics. Any function used to represent learning must consist of two distinct parts: a *response variable* and a behaviourally or cognitively

defined *state variable*. The response variable expresses the observed change in the behaviour of the system. The state variable expresses the specific or non-specific changes in the disposition of the system to learn. In the case of non-specific change in disposition, learning is promoted and persists over an extensive interval of time. This is referred to as the formation of a learning set, or as learning to learn (e.g., Harlow, 1949). Conversely, non-specific transfer effect that is increasingly negative and inhibits learning is called fatigue.

Similar descriptions of learning surface in the generic mechanism at play in the acquisition of an artificial grammar. This can be described as a modulation over time of a set of transitional probabilities that are position sensitive in a sequence of events (e.g., in a given sequence, what is the likelihood of *Letter(t+1)* given *Letter(t)*). In the starting state we regard all probabilities to be equal. This is an approximation and the true value depends on previous acquisition routes through similar experiences. During acquisition the cognitive system will reinforce connections related to those conditionalities that get imprinted through repeated exposure of regular patterns in the environment (i.e., the environmental input consisting of the acquisition sequences acting as the regular patterns).

1.3.5 Learning paradigms

Different memory systems may require different learning modes. Learning by instruction or supervised learning implies a rich source of external feedback—a teacher. An example would be error-based learning paradigms provided with detailed directional information which is utilized (e.g., error back-propagation) by the learning system to improve performance (Arbib, 2003; Haykin, 1994). Another example is the trial-and-error based learning or reinforcement learning, which is a weaker form of environmental interaction still dependent on external feedback. In this type of learning a (cognitive) system learns through trial-and-error interaction with the environment to gradually select appropriate actions by being provided external feedback in the form of reward signals. This can be conceptualized as searching through the available model space. The strategy introduces an important trade-off in the form of an exploration-exploitation dilemma, i.e. how the learning system should allocate its temporal resources given a finite life-time. With respect to exploration: how much time the system should spend attempting improve its model of the environment with the objective of optimizing exploitation opportunities (i.e., performance). With respect to exploitation: how much time should be devoted to utilizing what has already been acquired in order to achieve the primary objectives of the learning system in the first place (cf., Sutton & Barto, 1998). Reinforcement learning represents a type of guided learning where positive and/or negative feedback is provided based on the outcome of an action without any detailed directional

information. The learning system is not instructed how to change its internal workings but only an evaluation of whether a certain choice of action was in some sense appropriate or not.

Learning and adaptation can, as an alternative to supervised learning, take place without any external feedback, by means of unsupervised or self-organised learning. This implies that the outcome of the acquisition process is determined by the interaction between the input experienced and prior structure as well as properties of the learning system. A self-organised learning process may structure a neural network to represent the type of environmental structure it encounters and the adaptive process is sensitive to (e.g., correlation structures in the environment, cf., Rieke, Warland, van Steveninck, & Bialek, 1996). Another unsupervised learning example is based on internal monitoring of system performance based on internal measures of error or consistency. The internal measures can improve internal processing pathways, such that for example parts of the nervous system monitor other parts providing adequate teaching feedback.

To understand neural processing within an evolutionary framework one have to consider that the human brain has an evolutionary history on the order of 1 billion years (Deacon, 1997), and that a neural system controls behaviour with local and global consequences in terms of survival and reproductive success. Evolvability, the property of a genetic system to tolerate mutations and modify the genotype without seriously reducing its phenotypic fitness, must have provided, it seems, a selective advantage. Koch and Laurent (1999) suggest that the property of evolvability favoured compartmentalisation/modularity, redundancy, weak and multiple parallel linkages between regulatory processes as well as component robustness (for a different perspective, see Fodor, 2000). Sufficient stability and tolerance for evolutionary modification is provided if several of the constituent components and their coupling links are not crucial for survival but can serve as a substrate for the evolutionary search for fitness. Indirect evolutionary pressures lead to neural systems replete with specialized circuits, parallel pathways, and redundant mechanisms, with the effects of neurobiological evolution as a mechanism for the incorporation of prior structure into the processing infrastructure.

1.3.6 Cognitive models

Historically cognition has been conceptualized in terms of on the one hand the classical cognitive rule-based symbolic processing paradigm (Fodor & Pylyshyn, 1990) or on the other hand in terms of parallel sub-symbolic processing at a certain level of abstraction (Shastri & Ajjanagadde, 1993; Smolensky, 1988). Regardless of ones beliefs on the matter it is clear that

cognitive functions are implemented in the network architecture of the brain and depend on the processing characteristics of such networks.

When dissecting the classical cognitive paradigm it is important to realise the differences between the computer architecture and the human brain. The construction of logical gates in a computer is homogeneous and non-adaptive and the connectivity density of the gates is low compared that of the brain (cf. e.g., Savage, 1998; Tanenbaum, 1990), yet support memory on different time-scales and processing capacity. The logical gate in the central processing unit of a modern microprocessor is averagely connected to other gates on the order of 1-10, a connectivity factor 1000-10000 less than inter-neuronal convergence and divergence. The brains neural systems also wire themselves during ontogenetic development and its initial circuitry is up for modification from the environmental input throughout both the life of the individual as well as more short-term modifications driven by learning. The dynamical system of the brain function differ significantly in terms of information processing from present day computer architectures in the scale of structural and dynamic complexity. A computer derived from the classical Von Neumann Machine incorporates a general purpose central processing unit which exerts finite state control over the process flow (Minsky, 1967; von Neumann, 1957). The computer processing is globally synchronized through a clock-frequency or highly coordinated through just-in-time processing. Specifically challenging for cognitive neuroscience is the absence of global process coordination in neural systems.

Since the classical cognitive science perspective not easily translates into the processing characteristics of neural systems this could suggest a fundamental problem with either the classical view on cognition (e.g., Charniak, 1993; Churchland & Sejnowski, 1993; Edelman, 1990; Rumelhart & McClelland, 1986), or for the discipline of neuroscience rather than cognitive science per se (Chomsky, 2000). However, advances in non-classical information processing in dynamical systems promote the integration of the classical cognitive science framework within a general dynamical systems framework.

In the framework of classical cognitive science and artificial intelligence (cf. e.g., Charniak & McDermott, 1985; Fodor, 1983; Newell, 1992; Posner, 1989; Stillings, Weisler, Chase, Feinstein, Garfield, & Rissland, 1995) information is assumed to be coded by *data structures*, while cognitive processing is accomplished by algorithmic *rule* operations executed on the *symbol* representations which underlies the data structures. Such rule based symbolic processing paradigm (cf. e.g., Horgan & Tienson, 1996; Wilson & Keil, 1999) suggests that

cognitive phenomena can be modelled and that isomorphic models of cognition can be found within the framework of Church-Turing computability (cf. e.g., Cutland, 1980; Davis, Sigal, & Weyuker, 1994; Lewis & Papadimitriou, 1981; Rogers, 2002).

From this perspective, a cognitive system consist of a state space of internal states and computations are instantiated as transitions between states while optionally receiving input and generating output as determined by cognitive transition functions or non-deterministic transition relations, generating trajectories in state space (Cutland, 1980; Davis et al., 1994; Lewis & Papadimitriou, 1981; Savage, 1998). This implies that information processing can be reduced to the topology of state-space trajectories determined by the dynamical system.

For computations in the framework of Church-Turing from a time-discrete dynamical systems perspective, consider the case of a cognitive transition function with an input space, a space of internal states, and an output space, where all spaces are finite. A possible cognitive transition is determined or governed by the function $Transition: StateSpace \cdot InputSpace \rightarrow StateSpace \cdot OutputSpace$. Suppose that at processing step n the system receives $input(n)$ when in $state(n)$. This makes the system change state into $state(n+1)$ generating $output(n+1)$ according to (**Equation 1.2**)

$$(state(n+1), output(n+1)) = Transition(state(n), input(n)) \quad (1.2)$$

forcing the processing system to be driven by reading the input stream [..., input(n), input(n+1), ...] into a trajectory in a finite state space [..., state(n), state(n+1), ...] with the result that the system generates the output (alternatively, constructs the sequence of actions) [..., output(n), output(n+1), ...].

It is also needed to explicitly described the memory organisation of the computational system (cf., **Table 1.2**) since memory properties in terms of storage capacity (e.g., finite or infinite) and accessibility (e.g., stack- or random access) determine the computational power of the processing architecture (see e.g., Davis et al., 1994; Lewis & Papadimitriou, 1981; Savage, 1998).

Architecture	Machine complexity	Memory organisation
---------------------	---------------------------	----------------------------

		States	Registers	Stack	Accessibility
Finite-state	Finite	Finite			
Push-down	Finite	Finite		Infinite	Top of stack
Turing	Finite	Finite	Infinite		Random

Table 1.2. Outline of the Chomsky hierarchy and its respective architecture memory organisation (adapted from Petersson, 2005a).

Transitions between internal states during the process of receiving input, storing intermediate results of the computation in memory, and generating output, all subserve information processing. With respect to the mechanism underlying the transitions in state space there is no fundamental distinction in terms of machine complexity between the different computational architectures (**Table 1.2**) although there are differences in expressivity which are fundamentally related to the interaction between the generating mechanism and the available memory organisation. Machine complexity, in other words the complexity of the computational mechanism of the architecture, must be distinguished from memory organisation complexity. Whether the storage capacity of the memory organisation is finite or infinite is crucial for the level of nesting that can be expressed in the generated output, an important aspect of a systems expressivity. For all architectures, a *Transition* function can be realised as a finite-state machine (**Equation 1.2, Table 1.2**). This also holds for the control of the general purpose architecture of the universal Turing machine (cf., Hopcroft et al., 2000; Savage, 1998; Wells, 2005).

The important determinant of expressivity is the availability of infinite storage capacity. A memory organisation that allows for infinitely recursive processing capacities is able to realise functions of high complexity, and will achieve complex levels of expressivity (Cutland, 1980; Rogers, 2002). In language modelling in theoretical linguistics and psycholinguistics, among other cognitive domains (cf. e.g., Newell, 1990; Russel & Norvig, 1995), the classical framework has served well (cf. e.g., Partee, ter Meulen, & Wall, 1990; Sag, Wasow, & Bender, 2003). A fundamental hypothesis of generative grammar (Chomsky, 1957) is that it is possible to give an explicit recursive definition of natural language (or at least for syntax) and all commonly used formal language models can be described within the classical framework (Wasow, 1989).

1.3.7 Levels of description of cognitive processes

Cognitive models of information processing, formulated within the classical framework of cognitive science is best analyzed at three levels of description (Marr, 1982):

1. *The functional/computational level* specifying in formal terms the function that results from the processes of the system, i.e. a formal theory for the function computed by the system.
2. *The procedural/algorithmic level* which given a formal theory specifies the representations and procedures for processing these representations (i.e., *InputSpace*, *StateSpace*, *OutputSpace*, and *Transition: StateSpace • InputSpace → StateSpace • OutputSpace* in **Equation 1.2**)
3. *The implementational/hardware level* which given an algorithmic description specifies how the representations and procedures are implemented in a physical system.

Starting at the functional/computational descriptive level of a cognitive model, the architecture supporting the cognitive model is an abstract functional organisation to which the physical state transitions are systematically related, independent of any particular implementation. However, for models of cognition that claims to model physically realisable systems an important constraint is that any processing has to be possible to implement in a physical device (Horgan & Tienson, 1996). To be able to implement the cognitive model in a physical device, the cognitive function must be describable not only at a functional level, but simultaneously at a procedural and an implementational level. The procedural description has to be within reasonable computational complexity, logical depth and machine complexity, while the implementational description must meet finite memory constraints and real-time and space constraints of the physical device. This means, for example, that the only possible algorithms that can be implemented are those that consumes computational resources scaling at a low polynomial order with the problem size (e.g. 2) in time and space, and not exponentially (Hopcroft et al., 2000). Unconstrained models of natural language, such as syntax parsing and comprehension, might be problematic.

Classical cognitive science is also associated with the concept of modularity, in the sense that the cognitive architecture is thought as being divided into well-defined communicatively interactive sub-components. Modularity is connected with but not dependent upon genetically determined and informational encapsulated structures where these modules are viewed as

input-output devices, isolated from lateral or top-down influences between modules and feeding a central domain-general processing module (cf. the high-level feed-forward perspective on cognition in Fodor, 1983). In this way, cognition is commonly divided into functional domains, further divided into sub-domains and cognitive components/processes, as in for example language and the temporal organisation of behaviour.

The modular view of brain function is supported from neuropsychological lesion data, while the genetic specification of these cognitive modules is supported from data on developmental disorders, although these it can be interpreted differently (e.g., Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Paterson, Brown, Gsodl, Johnson, & Karmiloff-Smith, 1999; Plaut, 1995; Young, Hilgetag, & Scannell, 2000). The classical perspective of modularity is difficult to integrate with a neurophysiological perspective on brain function, given the current understanding of the coding and processing principles of the nervous system (cf. e.g., Arbib, 2003; Gerstner & Kistler, 2002).

Cognitive functions are subserved by interactive and parallel distributed processing principles. The brain organisation resembles a hierarchically structured recurrently connected network indicates that the structure-function relationship is complex. Each brain region may dynamically participate in several functional networks and may serve different functions depending on the functional context in which it operates at a given moment of processing. Information is believed to be represented as distributed activity and information processing is thought to emerge from the interactions between different specialized regions. The challenge in cognitive neuroscience is to work out how these functions arise from brain network architectures.

1.3.8 Memory systems

Human memory is composed of several memory systems which operate on different characteristic time-scales. These memory systems most likely interact more or less in different constellations of subsystems or processing architectures (Gabrieli, 1998; McClelland, 1994). Attentional processes interact intimately with certain learning processes. In this context cognitive processes can be divided into *controlled processes* requiring a high degree of attention and *automatic processes* requiring a low degree of attention, where automaticity seems to develop gradually from controlled to automatic behaviour. Performance of a novel task is thought to rely on a higher degree of attention and controlled

processing while performance become more automatic with practice requiring less attention and controlled processing.

One coarse division of memory along the dimension of time-scales is the division into *short-term* and *long-term memory*. One influential model of short-term memory is the Baddeley-Hitch model (Baddeley, 2000; Baddeley & Hitch, 1974). The Baddeley-Hitch model is actually a model of *working memory* since it combines short-term encoding, storage, and retrieval of information with components which are thought to support several higher cognitive functions, including reasoning and language (Baddeley, 1986). In this model, working memory consists of a central executive with three support systems. One component is the *phonological loop* which supports short-term encoding and storage of verbal information. A second is the *visuospatial sketch pad* supporting short-term encoding and storage of visuospatial information. The third component is the *episodic buffer* which is a capacity limited system that provides temporary storage of information held in a multimodal code. Its key function is integration between the different subcomponents of working memory and to provide an interface between the working memory components and long-term memory. It is thought to be capable of binding information from the subsidiary systems as well as from long-term memory into a temporary episodic representational format (Baddeley, 2000). The transient and early role of the medial temporal lobe system in long term memory formation and sequence encoding, in conjunction with the prefrontal cortex, makes these structures likely candidates for the episodic buffer feature of short-term integrative role and episodic format (Eichenbaum, 2000; Simons & Spiers, 2003). The phonological store is putatively implemented in the left inferior parietal region (Brodmann's area (BA) 39/40) together with parts of the superior temporal cortex (BA 22 Becker, MacAndrew, & Fiez, 1999; Paulesu, Frith, & Frackowiak, 1993), with the articulatory rehearsal process involving a left frontal circuit including Broca's region (BA 44) and pre-motor cortex (BA 6/44, Smith & Jonides, 1998, 1999).

Human long-term memory is also composed of different memory systems (Tulving & Schacter, 1994). There exists a consensus concerning the broad division of human memory into subsystems supporting *declarative memory* and subsystems that do not, sometimes referred to as *non-declarative memory* systems since they all share the feature of not being accessible for conscious inspection (**Figure 1.5**). Declarative memory supports the capacity to encode, store, and retrieve information of facts and events. Declarative learning is a learning processes that are accessible to subjective conscious analysis during both learning and retrieval processes. Processes that were once consciously accessible (that they were

expressible or declarative) during the period of learning, but became out of reach for conscious description or exploration after neural compilation of some kind, are also considered a variety of unconscious learning abilities. Since an initial conscious access to existing learning abilities means that they were dependent on the existence of conscious processes to begin with, this kind of unconscious processes have to be of an evolutionary late origin.

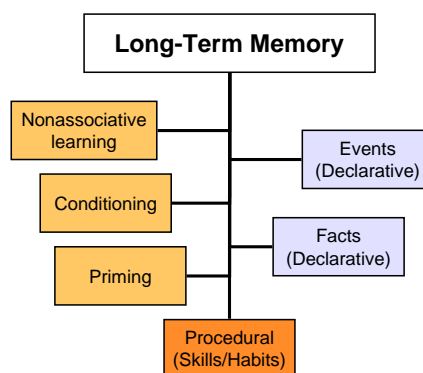


Figure 1.5 The taxonomy of the main types of human long-term memory systems.

Unconscious learning abilities which constitute a disjoint spectrum including everything from rudimentary neural processes to processes that involves networks of neural connections at longer distances, such as different forms of conditioning, habituation, repetition priming (e.g., facilitation of recognition, reproduction or biases in selection of stimuli that have recently been perceived), *incidental learning*, and *procedural learning* such as skills and habits (Knowlton, Mangels, & Squire, 1996a). These are all evolutionary old abilities shared within the vertebrate series and with many invertebrate species. They are fast and rigid processes that develop and are modified over longer periods of time during the development of the individual. The underlying neural implementations that subserve the different abilities span from single neuron receptor level to large-scale neural networks. The knowledge or information acquired by unconscious memory systems is commonly expressed through performance changes rather than explicit retrieval, and depends on the integrity of specific brain systems; for example the basal ganglia, the amygdala, and the cerebellum (Eichenbaum & Cohen, 2001).

A commonly used distinction of memory systems (Tulving, 1995), related and partly overlapping to the distinction between *declarative* and *non-declarative memory* systems, is that between *explicit* and *implicit memory*, where *explicit* and *implicit* usually refer to forms of memory expression. In this usage, implicit memory denotes the expression of memory without awareness of its acquisition or use; that is, behavioural expressions of what an individual has learnt without remembering how, when, or where the learning occurred (cf. non-declarative). In contrast, explicit memory commonly refers to the expression of what the individual is aware of and can explicitly report if probed (cf. declarative).

The typical example of procedural learning is skill acquisition such as master the art of riding a bike or master the motor control needed to pronounce words in what will turn out to be your native tongue, and yet being unable to put in words how this is achieved. Acquisition processes that are involved in the acquisition of procedural knowledge is an example of such non-declarative information acquisition, also called *implicit learning*. The outcome of implicit learning is referred to as *implicit knowledge*.

Investigations on different memory systems have mainly concentrated on different aspects of declarative types of memory but not that much on the unconscious variants such as procedural memory and mental skill acquisition, and more effort has been made to study procedural motor learning than non-motor learning (Seger, 1998). The focus of this thesis, on non-motoric procedural learning, is an example of investigations correcting this imbalance.

1.3.9 Implicit sequence learning

Implicit acquisition of knowledge about structured patterns embedded in stimuli can occur as an unintentional consequence of experience. This phenomenon can be found in for example the sensorimotor domain, language, and music (Stadler & Frensch, 1998). Seger (1994), following Reber (1967; 1989; 1993), suggested four characteristics for the phenomenon of implicit learning (Forkstam & Petersson, 2005):

1. Limited explicit accessibility to the acquired knowledge and subjects typically cannot provide sufficient (in many cases, any) explicit account of what they have learnt.
2. The nature of the knowledge acquired is more complex than simple associations or based on simple exemplar-specific frequency counts.

3. Implicit learning does not involve explicit hypothesis testing but is an incidental (automatic) consequence of the type and amount of processing performed on the stimuli.
4. Implicit learning does not rely on declarative memory mechanisms that engage the MTL memory system.

Thus, to characterize implicit learning it is necessary to address issues related to the nature of the acquisition process (e.g., implicit vs. explicit, automatic vs. controlled, incidental vs. intentional), the nature of the acquired knowledge and its representation (e.g., implicit vs. explicit access, abstract vs. concrete, structural vs. surface-based, complex vs. simple), and to characterize their functional role (e.g., implicit vs. explicit strategies, automatic vs. controlled processing, Dienes & Perner, 1999).

Implicit learning is typically investigated with three different stimulus structures (patterns, sequences, or functions) and three different response modalities (conceptual fluency, efficiency, or prediction and control, cf. Seger, 1994; Stadler & Frensch, 1998). Besides artificial grammar learning, one of the most intensely investigated implicit learning paradigms is the serial reaction time task, in which implicit learning is inferred from faster reaction times in responding to reoccurring vs. e.g. random sequences, while the participants typically report no or little awareness of reoccurring sequences. There are several proposals for how knowledge of sequence structure is acquired, including the acquisition of stimulus-stimulus, stimulus-response, response-response associations, or perhaps more abstract representations (cf. Stadler & Frensch, 1998). The learning of sequences with a fixed order can be viewed as a special case of acquiring knowledge about more general structural regularities or temporal contingencies in stimuli. These regularities can be deterministic, probabilistic, or non-deterministic.

In the serial reaction time task (Nissen & Bullemer, 1987) participants respond during a training phase to a target that appears on a computer screen in one of four horizontally arranged locations. Each location is mapped to a response key and participants are asked to respond as quickly and accurately as possible by pressing the key that corresponds to the current target location. Target locations on successive trials follow a systematic pattern that is continuously repeated throughout the training phase. Although participants are not informed of this sequential regularity, they nevertheless learn something about the deterministic structure of the task: When the systematic response sequence is replaced by random sequences response times increase, providing an indirect performance based measure of

sequence learning. Nissen and Bullemer (1987) observed reaction time savings for structured relative to random responses in six amnesic patients who reported to be completely unaware of a sequential regularity. This finding was confirmed in a study by Reber and Squire (1994) on amnesic patients showing normal sequence learning on the performance measure, while severely impaired in direct tests of conscious knowledge such as verbal report or recognition of the sequence.

Previous studies in serial reaction time task suggests that implicit motor sequence learning is sensitive to the statistical structure of sequences. Both absolute frequencies and transition probabilities are important, and these characteristics might be detected in different temporal order depending on the statistical properties of the stimuli (Lungu, Wachter, Liu, Willingham, & Ashe, 2004).

It is likely that that systems supporting both implicit and explicit learning can be engaged, and may perhaps also interact, during acquisition. Relevant stimulus dimensions (location/symbol) might engage different sequence learning mechanisms in the formation of internal representations (Tubau & Lopez-Moliner, 2004) suggesting a response-control shift (i.e., from stimulus-control to internal representational control) to correlate with the emergence of explicit knowledge in the symbol condition. Wilkinson and Shanks (2004) used a process dissociation procedure to separate automatic from intentional forms of processing (cf., Jacoby, 1991) and argued that sequence knowledge can be brought under intentional control.

It is well-established that skill-improvement not only occur during practice but also during off-line periods (i.e., between practice sessions), a so-called consolidation effect (for reviews see e.g., Maquet, 2001; Robertson, 2004; Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002). In a recent artificial grammar learning study, participants classified novel strings after a period with or without sleep (Nieuwenhuis, Folia, Forkstam, Jensen, & Petersson, In submission). It was found that sleep actively enhanced the extraction of grammar knowledge. The contribution of rapid eye movement (REM) and non-REM sleep in consolidation of implicitly acquired sequence knowledge is still not well-understood. In a serial reaction time task study, Robertson and colleagues (2004) showed a sleep-dependent dissociation for implicit and explicit sequence learning. Off-line improvement was sleep-dependent for explicit sequence learning and correlated with non-REM sleep, while off-line improvements in implicit learning seemed to be sleep-independent and only to depend on the time-interval

between practice-sessions. Cajochen and colleagues (2004) showed that the acquisition of different sequence structures improved after controlled sleep with multiple naps vs. sleep deprivation, in particular with naps that followed the circadian REM-peak.

The early findings reported in the implicit learning literature lent credibility to the position that memory is composed of functionally separate systems with distinctive neural structures (Cohen & Squire, 1980; Gabrieli, 1998; Schacter & Tulving, 1994a). In particular, the distinction between declarative and non-declarative memory (e.g. Squire & Zola, 1996) seemed to capture the critical finding well, that individuals can adapt to the statistical structure of their environments without being conscious of the underlying statistical contingencies. However, the multiple-systems view of conscious and non-conscious learning did not stand uncontested. An influential critique was formulated by Shanks and St. John (1994) concluding from an extensive literature review that the existence of dissociable conscious and non-conscious learning systems had not been established convincingly.

Their critique was primarily methodological. The standard demonstration of non-conscious learning requires dissociation between an indirect performance measure that indicates learning, and a direct test that indicates a lack of conscious knowledge. However, in order to accept this dissociation as evidence for non-conscious learning, one needs to presume that the direct test is sensitive enough to detect all conscious knowledge that might have been expressed on the performance measure (Reingold & Merikle, 1988). Empirical dissociations reported in the literature either did not withstand scrutiny or failed to replicate. For example, both Reed and Johnson (1994) and Destrebecqz and Cleeremans (2001) reported sequence learning on the indirect test and chance performance on a recognition test, but subsequent replication studies by Shanks and colleagues (Shanks & Johnstone, 1999; Shanks, Wilkinson, & Channon, 2003) provided no evidence of dissociation, keeping implicit learning an elusive phenomenon.

1.3.10 Knowledge representation and learnability

The perspective taken in this thesis views artificial grammar learning as a model for investigating implicit sequence learning (Forkstam & Petersson, 2005; Seger, 1994; Stadler & Frensch, 1998). This view holds that the capacity for generalisation that the subjects show in the grammaticality classification task is based on the implicit acquisition of regularities reflected in the input strings. Reber (1967) defined implicit learning as the process by which an individual comes to respond appropriately to the statistical structure in the input ensemble.

Also, he suggested that humans can acquire implicit knowledge of aspects of the underlying structure through an inductive statistical learning process and this is put to use during grammaticality classification. At the time, Reber (1967) argued that implicit learning mechanisms abstracted 'rule-based' knowledge, but has since modified his position (Reber, 1993), and more recent studies seem to suggest that dual mechanisms might be at play (Meulemans & Van der Linden, 1997).

In artificial grammar learning, the implicitness of the to-be-learned information can be understood in essentially two different ways. Even though the presented material in each situation is fully accessible during learning, in grammaticality classification the subject is informed and aware of the existence of a common underlying structure of the stimuli. In this situation the task is explicit to the subject (*classify grammatical from non-grammatical items*). In preference classification, the subject is not informed or aware of the existence of a common underlying structure of the stimuli. In this situation the task is implicit to the subject (*declare if you like or not the following items*). Importantly, in both these situations the structure to be learned is always presented implicitly, that is, the rules of the grammar is not specified but is instead represented in the grammatical items presented during acquisition. Furthermore, since the subjects are always tested on novel items never encountered before, in both these situations it is the capacity to learn the underlying structure that is tested rather than a mere memorization of exemplars on surface forms.

Support for the implicit character of artificial grammar learning comes from lesion studies of amnesic patients. Knowlton and Squire (1996) investigated amnesic patients and normal controls on a classical and a transfer version of the artificial grammar learning task, in which the subjects classified grammatical and non-grammatical items composed from a novel letter alphabet. The patients and their normal controls performed similarly on both artificial grammar learning tasks while the amnesic patients showed no explicit recollection of whole-item or fragment (chunks of two or three letters) information. Knowlton and Squire (1996) argued that these results indicate that the explicit recollection in the normal controls reflects an epiphenomenon not necessary for adequate performance on the classification task. Instead, they argued, artificial grammar learning depends on the implicit acquisition of both abstract and exemplar-specific information, the latter suggesting acquisition of distributional information of local sequential regularities. Furthermore, they argued for the existence of abstract representations (i.e., rule based representations) based on the performance of the amnesic patients and their normal controls on the transfer version of the artificial grammar learning task. It appears that humans are able to transfer knowledge acquired from exemplars

in one domain to a different domain (Gomez & Schvaneveldt, 1994). Similarly, Skosnik et al. (2002) suggest that artificial grammar learning involve the non-conscious consolidation of complex rules. In addition, it has been shown that infants have some capacity to learn and generalize over local regularities. Studies in young infants indicate rapid (on the order of 2 – 10 min) ‘rule-abstraction’ (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), the learning of transition probabilities in artificial syllable sequences (Saffran, Aslin, & Newport, 1996), and artificial grammar learning (Gomez & Gerken, 1999). In the study of Gomez and Gerken (1999), infants also demonstrated transfer capacity, suggesting that they were abstracting beyond the transitional probabilities holding between particular items in the grammar. However, it is an issue under discussion whether transfer studies demonstrated ‘rule-based’ learning or not. It is not clear that this conclusion follows from the results on transfer tasks. Specifically, this depends on the assumption that transfer performance is critically dependent on abstract representations and it is unclear whether this is necessarily the case. Transfer performance is dependent on a mapping from the representation of the acquired knowledge to the new surface form, which by necessity has to be established during the initial phase of the transfer task. It is not clear that it is easier to generate such a mapping from an abstract knowledge representation in comparison to a surface based representation (Redington & Chater, 1996). For example, it has been suggested that transfer results could be explained by simple similarity judgements and knowledge of substring regularities (Redington & Chater, 2002; Redington & Chater, 1996). On the other hand, the results of Knowlton and Squire (1996) suggests an abstract representation and the observation that the classification performance did not correlated with ACS strength when the participants had reached the late acquisition phase. This indicates that some form of abstraction of grammatical structure takes place. In addition, learning of long distance dependencies has been demonstrated in sequence learning as well as in artificial grammar learning (Ellefsen & Christiansen, 2000; Poletiek, 2002; Uddén, Araújo, Ingvar, Hagoort, & Petersson, In submission). This suggests that induction cannot be explained entirely in terms of the acquisition of local sequential regularities (cf., Meulemans & Van der Linden, 1997). While Reber (1967) originally argued that the implicit learning process abstracted ‘rule-based’ knowledge (however, see Reber, 1993 for a modification of his position), these more recent studies indicate that dual mechanisms is engaged (for an alternative perspective see Channon, Shanks, Johnstone, Vakili, Chin, & Sinclair, 2002; Johnstone & Shanks, 2001). In summary, it is reasonably clear from these studies that distributional information of local sequential regularities are acquired and used in grammaticality classification. But since ACS is independently manipulated in relation to grammaticality, this parallel sensitivity to ACS cannot explain the classification performance.

It is in this context of interest to recall that no super-finite class of languages, including the class of regular languages, is learnable from positive examples alone without additional constraints on the specific learning paradigm. This is for example the case in the formal learning theory framework of Gold (1967). At a first glance this appears to exclude the possibility of learning an artificial grammar from positive examples. It has also been suggested that this is the case when statistical learning mechanisms (cf. e.g., Cherkassky & Mulier, 1998; Duda, Hart, & Stork, 2001; Vapnik, 1998) are employed (Nowak, Komarova, & Niyogi, 2002). In the classical Gold learning framework (1967; cf., Jain, Osherson, Royer, & Sharma, 1999) it was assumed that the learning system had to identify the target language exactly based on only positive examples (i.e., well-formed strings); in addition it was assumed that the learning system has access to an arbitrarily large number of examples, while issues related to computational complexity were ignored. However, already Gold (1967) noted that under suitable circumstances this (un)learnability paradox can be avoided. These may for example include the existence and effective use of explicit negative feedback, prior restrictions on the class of possible languages, or if there are prior restrictions on the possible language experiences that can occur, that is, prior restrictions on the characteristics of the possible language environments. Results in formal learning theory confirm Gold's (1967) suggestion that, if the class of possible languages is restricted, then it is possible to learn infinite languages in infinite classes of formal languages from positive examples (Shinohara, 1994; Shinohara & Arimura, 2000); see also Jain et al. (1999). It should be noted that these constraints are of a general kind and not necessarily 'language specific'. As noted by Scholz and Pullum (2002), there exists classes of formal languages rich enough to encompass the string-sets of human languages and at the same time being identifiable from a finite sequence of positive examples. The restrictions determining these classes are not linguistically specific but general in character. Furthermore, the acquisition task becomes potentially more tractable if there are additional structure in the input or if only probable approximate identification success is required (cf. e.g., Anthony & Bartlett (1999) for an outline of the probably approximately correct learning paradigm and Engel & Van den Broeck (2001) for an alternative perspective).

Gold (1967) also suggested that the acquisition of super-finite classes of languages may be possible given reasonable probabilistic assumptions concerning the language environment and the initial language experience of children source (e.g., a stationary language environment on a suitable time-scale). Similarly, Chomsky (1981) suggested that negative evidence could be available without explicit corrections based on expectations (cf., Rohde & Plaut, 1999). One possibility is to generate expectations or predictions based on an internal

model. If the learning system has access to or can acquire a forward model, this can be used for model dependent prediction. This entails the possibility of an unsupervised learning framework in which error information (= difference[input, prediction]) drives the learning process (i.e., adaptation of the learning parameters driven by error information, see e.g., Duda et al., 2001; Haykin, 1998; Manning & Schütze, 2000). Simple examples of this is predictive adaptive time series models (Haykin, 1998) and predictive simple recurrent neural networks (e.g., Elman, 1990; Haykin, 1998). Connectionist modelling suggest that this is a viable approach to limited recursion (for a general overview see Christiansen & Chater, 2001; Christiansen & Chater, 1999; Seidenberg, MacDonald, & Saffran, 2002; see also Seidenberg, 1997). Simple recurrent networks may be viewed as a time-discrete analogue version of the finite state architecture (i.e., if real number processing is employed). It should be noted that simulations of a simple recurrent neural network, using finite precision numbers, effectively becomes a simulation of a finite state architecture. In summary, as noted by Scholz and Pullum (2002), it may be the case that language learning is not adequately described as ‘the logical problem of language acquisition’. Instead, formal learning theory (Jain et al., 1999) hold open the possibility that language classes of interest, at least in theory, can be acquired from weak environmental input consisting of a finite sequence of positive example (Pullum & Scholz, 2002; Scholz & Pullum, 2002).

In all included studies of the thesis we used a regular grammar, the simplest form of phrase structure grammar, which can be implemented as a finite state machine. It is commonly held that the class of finite state machines represents a restrictive class of computational models. However, from the point of view of machine complexity (e.g., the minimal size and depth measured in the equivalent number of logical operations or logical gates necessary to realise a given generative mechanism, cf. Savage, 1998), universal computational architectures like unlimited register machines (Cutland, 1980) and Turing machines (Davis et al., 1994) are, like finite state machines, finite in this sense. This can be understood by realising that the central processing unit of the register machine and the control unit of a Turing machine are examples of finite state machines (see Savage, 1998 and ; Tanenbaum, 1990, for concrete examples). The difference in formal language expressivity between regular and non-regular grammars (cf. Davis et al., 1994; Partee et al., 1990) springs necessarily from the memory characteristics of the computational system. In particular, formal language expressivity depends on the interaction between the computational mechanisms and factors like memory access (e.g., stack or random access) and most crucially on the memory capacity, that is whether this is finite or infinite (cf. Minsky, 1967; Savage, 1998). In a fundamental sense, it is the characteristics of the memory organisation that allow the computational architectures to

re-use their processing capacities (i.e., computational mechanisms) recursively to generate structurally rich languages (i.e., high expressivity). The computational machinery of universal architectures is no more powerful on its own than that of finite state machines, if finite memory constraints are imposed. The finite state machine is the only computational architecture of the Chomsky hierarchy (Davis et al., 1994; Partee et al., 1990; see also Savage, 1998) of infinite expressivity with respect to its fundamental recursive concatenation (i.e. to all its syntactic constructions), that is both finite with respect to its computational machinery and to its memory organisation. In addition, it is possible to implement finite recursion of any type in a finite state machine. From a neurobiological and cognitive neuroscience perspective it seems reasonable to assume that the human brain instantiate a finite storage capacity, both with respect to short-term working as well as long-term memory. This might highlight the importance of the neurobiological analogue of the finite state architecture. Interestingly, Hauser, Chomsky, and Fitch (2002) suggested that many of the important characteristics of the faculty of language are determined by neural computational constraints.

1.4 THE LANGUAGE FACULTY

The fact that all normal children acquire essentially comparable grammars of great complexity with remarkable rapidity suggests that human beings are somehow specially [innately] designed to do this.

(Chomsky, 1959, Review of Skinner's Verbal Behavior, p. 57)

The possibility that a finite-state grammar manage to model (or let say imitate) a context-free (centre-embedded; bracketed) grammar up to any finite level of embeddings by counting, urge any believer in the methodology of Ockham to shave off the excessive need of infinite context-free grammar in any cognitive task such as language.

An investigative approach to language acquisition have to take into account the necessity for each human being to use hypothesis testing at some level or another as the main approach to internalize a communicative language. Linguistics offers no choice to the behaviouristic approach. Everyone learn their language by the observation of human verbal behaviour, and through the observation of its own verbal behaviour amplified or corrected by the environment. There is nothing in a linguistic perspective beyond what can be collected from the open behaviour in observable situations (Quine, 1992). The artificial grammar learning paradigm represents a minimalistic model for this insight when investigating language acquisition and syntax processing in the laboratory. Even though language in communication is not easily investigated in artificial grammar learning, the paradigm captures something very essential in the sense of investigating the actual syntactic signal.

1.4.1 Objective 2

We will through investigation on artificial grammar learning validate and explore the repeated artificial grammar learning paradigm as a laboratory model to investigate the acquisition and processing of structural aspects of language, e.g. (morpho-) syntax processing.

1.4.2 Language acquisition

Some syntax related language mechanisms seem to be specific to the language function, while others might be shared with other psychological abilities, as well as with non-human co-species. As part of the broad language faculty, that is, the faculty of language in its

entirety together with parts shared with other psychological abilities, syntax functions as a regulator: it helps determine how the meanings of words are combined into the meanings of phrases and sentences. As a subset of the broader faculty, some aspects of syntax is included as part of the narrow language faculty, that is, aspects of language that are special to the language function (Hauser et al., 2002). Of these two levels of the language faculty, the latter is suggested to be typically human.

The function of syntactic structures in language is to code for who did what to whom, how, when, and where. Syntactic competence involves playing with basic who-does-what-to-whom categories such as Agent, Themes and Goal (Chomsky, 1981). Thematic roles such as Theme, Agent, and Goal bridge between semantic content in the concept structure and their syntactic relevance as structure regulators. Such features have the potential to have been prime targets for the emergence of syntactic signal decomposition (Bickerton, 2000). In this process syntax decompose a message or signal into a structure that is predictable for the recipient to receive, which lowers the energy cost of communication. This syntactic signal decomposition might have co-evolved between the sender and recipient within the given species (e.g., humans) as a mean to cut the communication costs. The delimiting mental resource of working memory would in this way be relieved. For the sender the message produced and sent could be of a lower quality as well as not well formulated, and for the receiver the predictive structure in the message or signal would free mental resources for parallel analysis of the underlying intension of the sender.

The general combinatorial process employed during syntax processing are the hierarchical order of phonemes and morphemes into words, and words into syntactic phrases, which are then collapsed and flattened into a serial order of words and phrases within phrases. In the opposite direction a sequential input of phrases is transformed into a hierarchical order of words and phrases. Certain local and (non-adjacent) long-distance dependencies are involved in this process, by setting constrain to the flexibility of order a given language can allow. Important examples of local dependencies are agreement and case marking. In agreement dependencies, verbs or adjectives are marked with inflections that correspond to the classificatory features of syntactically related nouns such as number, person, and grammatical gender. In case marking dependencies, noun phrases are marked with inflections, such as nominative and accusative, depending on the grammatical role of the phrase with respect to a verb, preposition, or another noun. Long-distance dependencies can on the other hand allow a question word or a relative pronoun to relate to distant verb (Pinker & Jackendoff, 2005).

The principles of statistical learning, a fundamental mechanism of information processing in the human brain, contribute significantly to language acquisition. Empirical data suggests that the process of word segmentation, acquisition of a lexicon, and acquisition of simple grammatical rules can be entirely explained through statistical learning. Statistical learning is mediated by changes in synaptic weights in neuronal networks and its mechanism constitutes a direct bridge between behaviour and the molecular biology and neurophysiology of the neuronal synapse. Even if it is unlikely that all aspects of language acquisition can be explained through statistical learning, some principles of effective language training are obvious already. Most important is the massive, repeated interactive exposure. Conscious processing of the stimulus material may not be essential. Instead, the crucial principle is a high co-occurrence of language and corresponding sensory processes engaged with the analysis of the environmental input (Breitenstein & Knecht, 2003).

1.4.3 The language acquisition device

The acquisition of our first language is commonly assumed to take place largely implicitly and almost entirely from unlabelled positive information alone (i.e., almost without explicit feedback as well as negative evidence). The likelihood that a child would manage to acquire a language in all its complexity from an environmental input of such sparsity has made some researcher to assume the task to be impossible unless the child has an inborn knowledge of the possible grammar of the language to acquire. Thus, it has been argued that the capacity of children to acquire natural languages depends on an innate universal grammar that constrains the form of possible human languages (Chomsky, 1965, 1986; Chomsky & Lasnik, 1995). The concept of rules and grammatical constructions are entirely rejected to give way for general interacting linguistic constraints from which the properties of linguistic expressions are derived. These principles are thought to be not learned or acquired, but innate. This inborn knowledge, which instantiates the universal principles and parameters of human languages, is supposed to be subserved with a dedicated language function called the language acquisition device. The variation between natural languages is accounted for by different parameter settings which set during the acquisition process (Jackendoff, 2002; Kager, 1999; Radford, 1997, 2000).

Although several researchers might believe differently, that language is not under natural selection pressure but instead is a perfect organ, such a statement is not congruent with the evolutionary theory. This postulates that natural selection also affect the language acquisition device. The process of signal decomposition instantiated as syntax processing in natural language is under phylogenetical pressure in the shape of evolution and natural selection, as

well as under ontogenetical pressure in the shape of individual development. The ontogenetical force set and tunes the syntax processing function within individuals to the communication and language space of the surrounding, including between generations. The phylogenetical force shapes the initial state of the processing architecture to possess capacity for neural process encapsulation which can be functional in syntactic analysis.

A phylogenetical and ontogenetical potential precursor to protolanguage is the mechanism of imitation, or mimesis (Vihman & DePaolis, 2000). Whatever neurological structures that support mimetical memory, e.g. gestural memory or memory for sequences, those structures may constitute the phylogenetic precursors to those that support linguistic representations in humans today. The phonological loop, one of the slave systems of the working memory system, provides the entry point to long-term memory storage for phonetic strings and larger phonological units composed of phonetic strings. Two observations suggest that this device may be more closely related phylogenetically to the processing of temporal sequences of meaningful gestures than to the processing of non-speech auditory patterns. The phonological loop has a critical role in motor production through subvocal rehearsal (Baddeley, 1986; Baddeley, Gathercole, & Papagno, 1998), and the processing of sign language in native signers depends on the same brain structures as the processing of speech (Poizner, Bellugi, & Klima, 1991).

Working from the perspective that human language evolved out of bodily actions in the shape of primate gestural systems in our evolutionary past (Arbib & Rizzolatti, 1997; Corballis, 2003; Rizzolatti & Arbib, 1998), an ever-present form of nonverbal action might have developed in the shape of a co-speech hand gesture (Kelly, Ward, Creigh, & Bartolotti, 2007). In its present-day role hand gestures together with spoken language would then impact speech comprehension in the brain, an influence fundamentally automatic in nature. This implies a specific relationship between gesture and speech within the domain of learning a first and maybe also a second language where gesture may act as a powerful tool in inducing neural change associated with the learning of a novel vocabulary. By not looking at language in isolation but instead putting it back in the body a more complete and accurate understanding of how people comprehend, develop and learn language in everyday communicative contexts might be generated. Hence, syntactic sequence processing might have its precursor in motor sequence control working with vocal soundings in its output, a view compatible with the ontogeny of the human child.

From the partial function of syntactic structure as to cut the costs of communication, it is interesting to imagine a version of protolanguage that possibly evolved within the primate lineage with vocal calls and body, hand and facial signing. Communication without syntax would be slow and displaced from real time reference. Thus an evolutionary pressure for nature to select for individuals with higher working memory capacity would be instantiated. As the output of the signal turned more acoustic/vocal an additional selective advantage would have been the case for individuals capable to comprehend and produce regulated message short cuts, possibly through means of calculating predictive feature of the signal detected by the receiver in an implicit fashion, in combination with the real time analysis of the underlying intention of the sender (instantiated as the mental capacity of Theory Of Mind). The interface for such a calculation would hook on to and unite nodes of mental capacity for sequentially controlled body posture, facial, manual, and vocal motor skills as well as visual and acoustic (implicit) feature detection. The suggestion of ventrolateral cortical regions sensitive to syntactic regulations thus seem comprehensible as target regions for further detailed analysis of involvement in syntactic feature calculations. Surrounding regions are receptive of phonological (BA 44/45/9), manual motor skills (BA 6/9), working memory (BA 46), and conceptual knowledge (BA 44/45/47). It is connected through reciprocal neural projections to basal ganglia (caudate) and thalamic structures sensitive to sensory structure features.

1.4.4 Formal grammars

The definition of a formal language includes a finite lexicon (alphabet) V of terminal symbols, $V = p$. The set of all possible finite symbol strings that can be generated from the alphabet V is given by Kleene-star operator $V^* = \{\emptyset, t_1, \dots, t_m, t_1 t_1, \dots, t_n t_m, \dots\}$. A formal language L over V is then defined as a subset of V^* , $L \subseteq V^*$; where a symbol string $s = t_{k1} \dots t_{km}$ is well-formed or grammatical if and only if $s \in L$ (cf. e.g., Hopcroft et al., 2000). This description of formal language emphasise an extensional or E-language, where the language is identified with its string set. This is adequate for formal investigations, but of limited meaning in the context of natural language grammars (Chomsky, 1986, 2000). A more fruitful approach from a cognitive point of view takes as its point of departure an intentional definition, an I-language (cf. Chomsky, 1995; Chomsky & Lasnik, 1995). This entails the specification of a generating mechanism, including principles of combinations and additional non-terminal symbols, capable of generating all grammatical (well-formed) strings and only those in a given language (e.g., Davis et al., 1994; Lewis & Papadimitriou, 1981; Taylor & Taylor, 1997). The generating mechanism serves as an intentional definition of the language, an I-language, and a string s is grammatical ($s \in L$) if and only if the formal mechanism (or

machine) can generate it. The term ‘language’ in formal language, does not entail anything beyond what is outlined above. A formal (or artificial) grammar represents a specification of a mechanism that generates (or recognizes) certain types of structural regularities.

A non-trivial yet simple formal model incorporating the idea of the infinite use of finite means is represented by a family of artificial grammars called regular grammars (finite state grammars) (Davis et al., 1994; Lewis & Papadimitriou, 1981; Partee et al., 1990). A finite state grammar is defined by:

1. A finite set of internal states $S = \{S_1, \dots, S_i, \dots, S_m\}$,
2. A vocabulary/alphabet of input/output terminal symbols $V = \{t_1, \dots, t_k, \dots, t_n\}$
3. A finite set of transitions $T = \{(t_1, S_1), \dots, (t_k, S_i), \dots, (t_n, S_m)\}$, a machine table describing the possible trajectories in state space of the automaton from its internal state S_i to S_j during which it will output the terminal symbol t_k .

Different generating mechanisms have different expressive power in terms of the languages they generate (cf. **Table 1.2**). The finite state automaton corresponds to the class of regular languages, which includes finite and some types of infinite languages, the latter examples implying that the general finite state automaton exemplifies a generating mechanism for the infinite use of finite means implementing a specific type of recursion. A regular grammar (**Figure 1.1**) generates right-branching phrase structure trees and the class of regular grammars has a generative capacity equivalent to the class of regular languages and can be implemented in the finite state machine architecture (see for example, Davis et al., 1994; Taylor & Taylor, 1997). A finite state machine can be viewed either as a language generator (Chomsky, 1957; Lasnik, 2000) or language recognizer for a given regular language (Cohen, 1997; Davis et al., 1994; Savage, 1998). Developments in mainstream generative grammar suggest that two important processing devices capture human syntactic competence, namely, *merge* and *move* (Chomsky, 1995; Radford, 1997) and it is possible to further reduce *move* to *copy + merge* (Berwick, 1998; Epstein, 1999). The family of right linear phrase structure grammars is closely related to and can be implemented by a constrained Merge operator. They are equally easy to implement within the framework of unification grammars (Jackendoff, 1997; Shieber, 1986) by for example a constrained unification operation (Vosse & Kempen, 2000), or within most commonly formal approaches to grammar (Sag et al., 2003).

Under certain circumstances, no super-finite class of languages are learnable from positive examples alone (Gold, 1967) even when powerful statistical learning mechanisms are put to use (Nowak et al., 2002; Vapnik, 1998). This includes regular, context-free, context-sensitive, recursively enumerable languages, etc. Under suitable circumstances this unlearnability paradox can be escaped (Gold, 1967), e.g. by explicit negative feedback, or if there are a priori restrictions on the class of possible natural languages, or on the possible initial language experiences that can occur (Rohde & Plaut, 1999). This means that learning of super-finite languages may be possible given some reasonable probabilistic assumptions concerning the language environment and the initial language experience of children source (Gold, 1967). Further, if negative evidence is available without explicit corrections based on expectations, unsupervised learning could be possible in which error information drives the learning process such that the adaptation of learning parameters in the model are driven by the error information (cf., predicative simple recurrent neural networks, e.g. Haykin, 1998).

1.4.5 The Reber grammar

Formal grammars instantiate a descriptive specifications of a finite generating/parsing mechanism for a given language. A formal grammar brings an explicit description of grammatical complexity. The Reber grammar (**Figure 1.1**) is an example of a formal grammar. The Reber grammar is a right-linear phrase structure grammar, which can be implemented in a finite state architecture. This Reber machine is an explicit generating mechanism for a formal regular language (e.g., Davis et al., 1994), in the present case corresponding to the Reber language. It should also be noted that the finite state machine corresponding to a regular language (or a right-linear phrase structure grammar), can serve equally well as a language recognition device or as a generative mechanism. Thus, the Reber grammar serves as an intentional definition (Chomsky, 1986) of the Reber language, as opposed to an extensional definition (i.e., the string set of the Reber language), and represents a formal specification of a mechanism that generates and recognizes certain types of structural regularities.

In the cognitive architecture of language processing, the syntax parser is a module that attempts to generate structural descriptions (e.g., the parse trees) corresponding to a given input sentence. Formal grammars and their corresponding languages and machines, does not have to be interpreted within the domain of language. Instead, they are relevant for any cognitive domain which utilizes structured representations and processes operating on these representations: goal directed action planning and temporal organisation of behaviour,

complex problem solving, perception and generation of sound patterns in music, etc. (cf., Petersson et al., 2004).

The finite state mechanism generating the Reber grammar consists of a finite set of internal states with the non-terminal symbols {State 0, State 1, State 2, State 3, State 4, State 5, State 6}, an alphabet with terminal symbols {<, M, V, X, S, R, >}, and a finite set of transition functions or phrase structure rules {State 0 → # State 1; State 1 → M State 2; State 1 → V State 2; State 2 → S State 2; ...}.

The way a formal grammar is represented can simplify the detection of ungrammaticalities in language sequences. Following the tree adjoining grammar formulation (Joshi & Schabes, 1997), a formal grammar can be described as a set primitive structures, so-called treelets (Figure 1.6). A treelet consists of a root node connected to a foot node and terminal symbol. In Figure 1.6 for example, we have extrapolated the treelet from the transition graph representation of the Reber grammar that connects the transition from state 1 to state 3 in the finite-state machine.

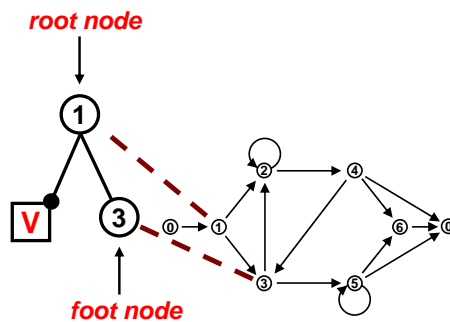


Figure 1.6 Tree adjoining formulation.

This process can be continued until we have transferred the complete structure representation of the Reber grammar to its corresponding treelets (Figure 1.7).

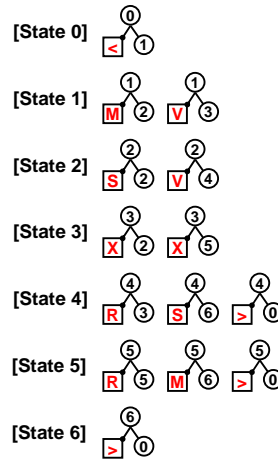


Figure 1.7 Structured representations of the full Reber lexicon expressed as treelets (Joshi & Schabes, 1997).

Primitives derived from the acquisition of a formal grammar are thought to be retrieved into the cognitive system during comprehension/parsing, and in the opposite direction, to be put together into larger complex during production. These processes are suggested to be performed in the human brain in a dedicated part of working memory, called the unification space (Hagoort, 2004). For example, as visualised in **Figure 1.8**, when encountered with the target item <MSSVRX>, the primitive composition of the sequence will be parsed in the unification space as a valid structure (see e.g., Forkstam et al., 2006). Learning in a cognitive system can in this way be viewed as expanding the nodes in a finite state grammar. For example, in phoneme acquisition, the division a phoneme into two separated phonemes is represented by splitting a node in two parts. Although the expected use of brain machinery might be more expensive, this visualisation of acquisition might generate a more adequate description of cognitive system acquisition.

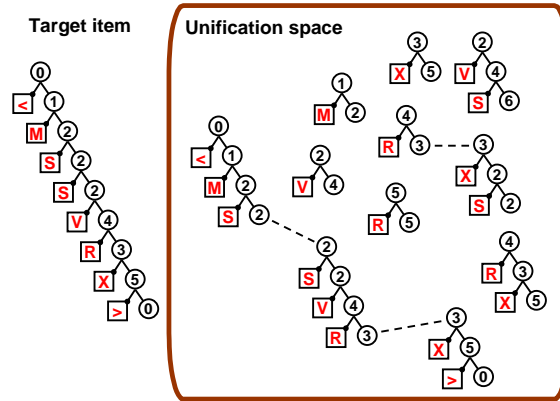


Figure 1.8 Sketch of the unification space. Primitives derived from the acquisition of a formal grammar are thought to be accessible to the cognitive system in the unification space. When encountered with the target item <MSSVRX>, the primitive composition of the sequence will be parsed in the unification space as a valid structure (see e.g., Forkstam et al., 2006; Hagoort, 2004).

1.4.6 Acquisition of formal grammars and artificial grammar learning

Artificial grammar learning is commonly conceptualized in terms of either a structure-based ('rule') acquisition or a statistical ('chunk-based') learning mechanism. An alternative view that is placed somewhere between these two more common conceptualisations, re-traces a major trend in theoretical linguistics since the early 80's, so-called lexicalisation (cf., Culicover & Jackendoff, 2005; Jackendoff, 2002), in which the distinction between lexical items and grammatical rules is vanishing and in effect shifts more of the grammar into the mental lexicon. In a *lexicalized* picture of artificial grammar learning, taking advantage of the fact that hierarchical structured information can be represented in terms of bracketed expressions (see e.g., Davis et al., 1994), the acquisition of simple structured representations (e.g., $[s_p, T, s_k]$, where s_j and s_k are *syntactic features* and T a *surface feature*) is akin to lexical learning, and the integration of such representations takes place in working memory during parsing (e.g., $[s_p, T, s_k] + [s_p, R, s_q] \rightarrow [s_p, T - R, s_q]$ if and only if $s_k = s_p$) after being activated (retrieved) by for example an input string. On-line incremental integration thus implements a recursive construction of complex representations (successive merging/unification, Petersson et al., 2004) from more primitive structures stored in long-term memory. We suggest that the latter process is dependent on general integrative mechanism supported by the left inferior frontal region. There is thus no need for a specific 'rule' acquisition mechanism, because the parsing process utilizes general integration mechanisms already in place for merging or

unifying structured representations. Finally, the syntactic features of lexical items have acquired the functional role of control in this picture (cf., finite-state control, Minsky, 1967; Savage, 1998), which is thus distributed over a long-term memory representation ('mental lexicon') in terms of the control features that govern the integration process based on the selection information that is allowed to merge (for further discussion see Forkstam & Petersson, 2005; Petersson, 2005b).

1.5 NEURONAL IMPLEMENTATION

The brain of an individual develops following a genetic blue print. On the observation at the level of the brain, each individual can unfold differently depending on the specific circumstances in the environment. This setup has the evolutionary positive effect of mutation condensation, and the developmental positive effect of time-locking the development of specialized organs, such as brains. In the end, it is the interaction between genes, the individual organism (or vehicle or survival machine), and its environment that will shape the development of the organism. In this section we will give an introduction to the implementation level of implicit sequence learning (see **section 1.3.7**) including a description of the corticostriatal circuits and their implications in implicit sequence processing.

1.5.1 Objective 3

We will through investigation on artificial grammar learning further our understanding of the specific neural processing architecture which subserves the language faculty and its syntax processing ability, an implicit non-motor procedural learning ability present in the human cognitive system.

1.5.2 Structural and functional complexity of the brain

The structure of the human brain is complex. Macroscopically its 1.5 kg mass is composed of grey and white matter. The grey matter contains neuronal cell bodies as well as local and short distance neuronal interconnections. At the neocortex, the outer layer of the brain, it forms a convoluted surface described as gyri and sulci, a structure that maximizes the possible numbers cells and their local connections. Grey matter is also localized deep inside the brain as neuronal nuclei: the basal ganglia, the medial temporal cortex, the cerebellar cortex and nuclei, as well as other subcortical nuclei in the mesencephalon and brainstem. The white matter contains the long distance interconnections of the brain of both corticocortical and corticosubcortical origin, inter-hemispheric tracts, as well as sensory input and motor output fibre tracts (Nieuwenhuys, Voogd, & van Huijzen, 1978).

Microscopically, the about 10^{10} neurons that form the brain is each working as an individual processing units supporting and receiving on the average more than 10^3 axonal output connections and a similar number of dendritic and somatic input connections. The brain is composed of some hundreds of trillions of interconnections using many thousand kilometres of cabling (Koch & Laurent, 1999). The synapse connections bridging the neurons consist of a double convoluted cell membrane with a massive amount of large proteins working as

receptors and ion-channels embedded in the surface of both neurons. In each synapse, the complex matrix of synapse proteins controls the influx and outflow of signal molecules between the connected neurons.

The brain connectivity structure is weakly hierarchical and consists of functionally specialized brain regions recurrently connected in a network. Each region consists in turn of several kinds of processing elements and synaptic connections (Felleman & Van Essen, 1991; Shepherd & Koch, 1998). Each processing element is independent and there is no focus for process control, in contrast to the central processing unit of the Von Neumann architecture (Tanenbaum, 1990). The organisational principles for cognitive brain functions depend on distributed connectivity patterns between functionally specialized brain regions as well as functional segregation of interacting processing streams, where the dominant pattern of interconnectivity is recurrent (see e.g., Gazzaniga, 2000, 2004). This complexity in this connection system is unnecessary for a system based on linear, sequential, and hierarchical feed forward information transfer, but essential for network processors that support interactive recurrent distributed processing. This suggests that parallel interconnected distributed anatomical networks, characterized by recurrent interconnectivity and functional integration across cortical networks, are essential for the processing characteristics of the brain.

The behavioural and cognitive significance of a given brain region is determined by its connectivity pattern, its local architecture and input and output connections, and by its neuronal subtypes and their distribution of for example receptor types and ion channels. In addition, the uniform basic outline of the neocortical architecture suggests that the functional role of a given brain region might be determined by its place in the neocortical macro-circuitry. It is suggested that each cytoarchitectonic area has a unique pattern of input and output connectivity and a corresponding pattern of task dependent functional connectivity (Passingham, Stephan, & Kotter, 2002). The functional architecture is additionally determined by the mechanisms that enable the processing systems to incorporate adaptive changes, allowing the system to learn as a functional consequence of information processing. Since the hypotheses of network circuitry (McCulloch & Pitts, 1943; see also Minsky, 1967) and neuronal assembly (Hebb, 1949), information processing in neural systems has been suggested to be accomplished such that information is represented as distributed activity in the brain and that information processing, subserving complex cognitive functions, emerge from the interactions between different functionally specialized regions or neuronal groups.

Fundamentally, all subsequent approaches suggest that cognitive functions emerge from the global dynamics of interacting sub-networks.

It appears that the basic computational units of the brain and their interconnections are relatively slow and imprecise in relation to the real-time task demands on processing performance (Koch & Segev, 1998). This seems to be related to inherent processing limitations of neurons. However, at least some neurons and neural systems appear to perform at levels not too far off from what is physically possible given its input and hardware characteristics (Rieke et al., 1996). For this to be accomplished, it seems plausible that the brain processes information interactively in parallel, and that processing properties that are rapid, fault tolerant, and robust emerge from these processing principles. The complexity at multiple levels of structure as well as function suggests that continued reductionism is not likely to lead to a fundamental understanding of cognitive brain functions from a complex systems perspective (Koch & Laurent, 1999). Instead, the detailed investigation of the nervous system has to be complemented by investigations at several different system levels. At present, higher cognitive functions of the nervous system are commonly characterized in terms of large-scale/macroscale concepts that are relevant at a behavioural level. An important objective of cognitive and computational neuroscience is therefore to bridge between the properties that characterize neurons, or neuronal assemblies, and the processing units and processing principles that are subserved by neural networks and are relevant to cognition. The most crucial scientific question still lingers: which is the neural code and how is the functional descriptions translated into this code.

1.5.3 Frontal lobe function

The prefrontal cortex is important for higher cognition. It is involved in the working memory functions supporting on-line maintenance, monitoring, manipulation, and selection, as well as in attentional processing, decision making, problem solving, language, non-automatic and flexible cognition and behaviour selection, the temporal organisation of behaviour and the decomposition of task processing into goals and sub-goals (i.e., prioritized dynamic scheduling or planning of sub-tasks), etc. How the prefrontal cortex subserves these cognitive functions is not yet well understood and various suggestions has been put forward: that prefrontal regions are specialized for different cognitive functions independent (Petrides, 1995) or dependent (i.e., modality specific, cf. e.g., Fuster, 1995, 1997; Goldman-Rakic, 1988; Goldman-Rakic, 1998), while others emphasize its adaptive nature (Duncan, 2001; Miller & Cohen, 2001). The primate prefrontal cortex has been investigated at the neuronal level on a wide range of tasks, including for example categorisation, working memory, rule

learning, rule switching, and cross-modal integration (for reviews see e.g., Duncan, 2001; Duncan & Miller, 2002).

The response properties of prefrontal neurons are very adaptable and it seems that any given neuron can be driven by several different kinds of input. This might be a result of the dense recurrent connectivity within the prefrontal cortex and with many other neocortical and sub-cortical structures (Fuster, 1997; Mesulam, 2002; Stuss & Knight, 2002). Following the adaptive coding model of Duncan (2001), working memory, attention and cognitive control are subserved by common processing properties of prefrontal cortex neurons in combination with the adaptable nature of these neurons. This would allow the prefrontal cortex to represent task-relevant information and provide a temporary, task-specific, context-dependent working memory space. Duncan (2001) suggests that this working space serves as a mechanism for selective attention and control by selecting task-relevant inputs, represented in the posterior neocortical regions, for further elaborate processing or manipulation of the task-relevant information.

Miller and Cohen (2001) argue that the prefrontal cortex stores representations of task-specific rules, attention templates, and task relevant goals. In their view, an important role of the prefrontal cortex is to bias the activation of goal related representations that are stored, represented, and processed in the posterior neocortical regions. They propose that this form of guided or controlled activation of posterior representations is essential for rule acquisition as well as the acquisition of new information and behaviours. They suggest that repeated activation of the same processing pathways creates stronger associations between posterior representations (i.e., stronger connections between posterior representational regions), while at the same time, the role of the prefrontal cortex gradually diminishes in controlling posterior neocortical processing as this becomes increasingly automatic.

Fuster (1995; 1997) propose that the common role of the prefrontal cortex is to temporally organise goal-directed behaviour and that this global function can be analyzed in terms of working memory, attention and inhibitory control, and he outlined mechanisms for monitoring, short-term memory and attentional selection that prioritize goals and task appropriate behavioural sequencing. Temporal integration is achieved by the prefrontal cortex in interactions with posterior cortical regions, determined by the modalities of task-relevant sensory and motor information. Moreover, prefrontal representations and processing are recruited in non-automatic behaviour, while well-practiced tasks can be performed

relatively independently of the prefrontal cortex. Several other researchers have sketched similar ideas in terms of a global workspace for non-automatic cognitive processing (Cohen & Servan-Schreiber, 1992; Cohen, Dunbar, & McClelland, 1990; Dehaene, Kerszberg, & Changeux, 1998).

Similarly, it has been suggested that the prefrontal cortex serves as a working memory that keeps stimulus representations active for on-line processing (Fuster, 1995, 1997; Goldman-Rakic, 1988). In particular, it is proposed that the prefrontal cortex, being part of an integrated network of regions including temporal, parietal, and limbic, is involved in the representation of stimuli in their absence. This would allow the prefrontal cortex to guide behaviour responses through internal representations in the sense of cognitive states. Related views from investigations of language processing suggest that the prefrontal cortex is engaged in structural integration that serve to rapidly and selectively bring together information in posterior representational regions. Different linguistic representations (e.g., phonological, syntactic, semantic, and pragmatic) are activated in parallel and integrated in a prefrontal workspace where incremental unification takes place (cf. e.g., Forkstam et al., 2006; Hagoort, 2003; Petersson et al., 2004).

Several of the outlined perspectives so far are processing oriented, but also representational perspectives on prefrontal function have been put forward (cf., Miller & Cohen, 2001; Wood & Grafman, 2003). For example, in the structured event complex framework outlined by Wood and Grafman (2003), it is suggested that prefrontal cortex stores representations of knowledge in the form of so-called goal-oriented sets of events. These goal-oriented sets carries a schematic sequence structure and represents various forms of knowledge, like event features, event boundaries, social rules, thematic knowledge, concepts, as well as grammars. The different aspects of a structured event complex are represented independently and they are encoded and retrieved in an episodic format (Wood & Grafman, 2003).

1.5.4 The left inferior frontal region

The left inferior frontal region seem to have a broad role in cognition including musical syntax, absolute pitch perception, and human imitation (Marcus, Vouloumanos, & Sag, 2003). An important general problem for models imply that the Broca's region in the left inferior frontal region is specifically related to different aspects of language processing is that neither neuropsychological lesion studies nor functional neuroimaging data appear to support such a strong hypothesis (Caplan, 1992; Dronkers, 2000; Kaan & Swaab, 2002; Marcus et al.,

2003; Zurif, 1990, 1998). Aphasia research requires a re-evaluation of the classical interpretation of the structure-function relationship based on the apparent double dissociation between Wernicke's (traditionally associated with left temporoparietal lesions) and Broca's aphasia (traditionally associated with left middle-inferior frontal lesions). Kaan and Swaab (2002) point out that these patients do not completely lack syntax processing capacities and they also exhibit some semantic deficits, suggesting that the Broca's region is neither necessary nor sufficient to induce syntactic deficits. Instead, they suggest, Broca's aphasia can be interpreted as a processing deficit in contrast to a knowledge deficit, and may at least partly be understood in terms of difficulties with certain aspects of temporal processing and integration of information, or in terms of short-term memory capacities. This suggestion is consistent with functional neuroimaging data indicating an important role for the prefrontal cortex, including the left inferior frontal region, in both short-term working memory and long-term memory (Cabeza, Dolcos, Graham, & Nyberg, 2002; Fletcher & Henson, 2001; Nyberg, Forkstam, Petersson, Cabeza, & Ingvar, 2002; Nyberg, Marklund, Persson, Cabeza, Forkstam, Petersson, & Ingvar, 2003; Simons & Spiers, 2003). In addition, functional neuroimaging studies comparing syntactically complex and simple sentences can be interpreted in terms of memory load and integration/unification of information (cf., Kaan & Swaab, 2002).

A growing body of evidence from functional neuroimaging suggests an overlap in the processing of structural relations in language and music (for a review see Patel, 2003). There seems to be a considerable overlap between regions implicated in the perception/production of music and the perception/production of abstract sequences, including the left inferior frontal region (Janata & Grafton, 2003). The similarities between music and language have recently been stressed (Hauser & McDermott, 2003; Patel, 2003; Peretz & Coltheart, 2003; Trehub, 2003). It has been suggested that music is a human universal, that it like language, organises discrete elements into hierarchically structured sequences according to syntactic principles (Jackendoff, 2002; Lerdahl & Jackendoff, 1983; Patel, 2003). Patel (2003) suggests that the commonalities between structural processing in language and music can be understood in processing terms such that brain regions engaged in processing these commonalities provide the neural infrastructure for structural integration. It is thought that the neural infrastructure engaged in structural integration are processing regions that serve to rapidly and selectively bring low-activation items in representation regions up to the activation threshold needed for integration to take place. Similarly, Hagoort (2003) proposed a language integration workspace in which integration of various sources of linguistic information (phonological, syntactic, semantic/pragmatic) operate in parallel in a workspace

where incremental unification takes place. The left inferior frontal region is hypothesized to be the core region subserving this workspace. It is suggested that during parsing, lexically specified structures enter the unification space where cross-talk between different sources of information immediately can influence the integration process.

There is also a growing body of evidence suggesting that Broca's region is not the only region related to the processing of syntactic information. Other brain regions which have been related to syntax processing include the left superior anterior temporal lobe, the left middle and posterior parts of the superior and middle temporal gyri, as well as right hemisphere regions (Bookheimer, 2002; Friederici, 2002; Kaan & Swaab, 2002). It therefore is reasonable to suggest that natural language syntax processing, or more generally the faculty of language, is dependent on a functional network of multiple interacting brain regions, none which is uniquely involved in syntax processing. With respect to this perspective, one might suggest that particular brain regions, for example the prefrontal cortex are computationally or processing specific (e.g., detecting and recognizing structural regularities; interpreting, integrating or unifying hierarchical regularities, or recognizing dependencies between related elements) independently of particular content domains. Specific brain regions may genuinely participate in a range of tasks, including Broca's region, with specialized function emerging from unique configurations of domain-general mechanisms (Marcus et al., 2003).

Meta-analyses of functional neuroimaging studies of syntax processing (Bookheimer, 2002; Friederici, 2002; Indefrey, 2004) report that the most reliably replicable finding related to syntax parsing across imaging techniques, presentation modes, and experimental procedures, are localized to the left inferior frontal gyrus (Brodmann's areas (BA) 44 and 45), consistent with what is known from brain lesion data (Caplan, 1992; Caramazza & Zurif, 1976; Friederici, 2002; Zurif, 1990). The left inferior frontal region is part of the prefrontal cortex, which has been related to different aspects of language processing, including phonological, syntactic, semantic, pragmatic, as well as non-linguistic contextual information (Bookheimer, 2002; Duncan, 2001; Mesulam, 2002). The prefrontal cortex has also been related to different short-term working memory and long-term memory processes (Baddeley, 2003; Simons & Spiers, 2003). In Baddeley's model of working memory (Baddeley, 1992, 2003), the phonological loop has been associated with the left temporoparietal and left inferior frontal regions. It has been suggested that the phonological loop may have evolved to facilitate the acquisition of language and in support of this notion, its capacity appears to be a good predictor of second language learning (Baddeley, 2003; Baddeley et al., 1998). The prefrontal cortex has been investigated in several primate studies at the neuronal level in a wide range of

complex tasks, including categorisation, working memory, rule learning and rule switching, as well as cross-modal integration (Duncan, 2001; Duncan & Miller, 2002). The response properties of prefrontal neurons are highly adaptable and any given neuron can be driven by different kinds of input, perhaps through the dense interconnections that exist within the prefrontal cortex as well as reciprocal connections to a majority of cortical and subcortical structures (Fuster, 1997; Mesulam, 2002; Stuss & Knight, 2002).

1.5.5 Functional anatomy of the corticostriatal circuits

Essential brain structures for procedural skill learning include the recurrent circuits connecting the prefrontal cortices with the basal ganglia and the thalamic nuclei. This seems to hold true also in the context of the acquisition and processing of implicit procedural knowledge subserving the mechanisms of artificial grammar learning. We will therefore here present an overview of general outline of these frontostriatal circuits (cf. e.g., Binder, Hirokawa, & Windhorst, 2008; Gazzaniga, 2004; Haber & Gdowski, 2004; Hendelman, 2006; Wise, Murray, & Gerfen, 1996).

1.5.5.1 The basal ganglia

Buried deep within the cerebral hemispheres lies *the basal ganglia* (see **Figure 1.9**), a collection of interconnected subcortical structures in the telencephalon, diencephalon and mesencephalon, constituting an interface between these brain departments originating from the early divisions of the embryonic brain. Diseases affecting the basal ganglia lead to characteristic disturbances of movement and of resting muscle tone, indicating its involvement in movement control. Even though the most obvious symptoms of the basal ganglia are related to the motor system, clinical and experimental evidence indicates that the basal ganglia also play a role in higher mental functions. Indeed, fibre tracing methodology show that the main efferent connections of the basal ganglia do not descend to motor nuclei in the brain stem and spinal cord but are rather directed to various cortical areas including the prefrontal. The cerebral cortex also provides the main input to the basal ganglia. There are two main outputs, one descends to the brainstem motor systems and the other is directed toward the thalamus, which in turn projects back to cortex. The basic cortical basal ganglia loop is in its outline connecting the cortex-basal ganglia-thalamus-cortex.

The striatum constitute the main input interface of the basal ganglia. It comprises three nuclei, all similar with respect to cells types, general projections, transmitters and receptors: *the caudate nucleus*, *the putamen* and *the nucleus accumbens*. The nucleus accumbens joins with

the striatum, the collective term for the caudate nucleus and the putamen. The caudate nucleus is localized deep within the cerebral hemispheres of the telencephalon. Although the caudate nucleus is one structure, it is regularly referred to as *the head* (in the frontal lobe), *the body* (in the parietal lobe), and *the tail of the caudate* (in the temporal lobe). The caudate is histologically similar to the putamen and the both regions originate embryonically from the same neurons. The strands of tissue connecting the caudate and the putamen in the adult human brain are a developmental rest of the internal capsule, a white matter fibre system separating the caudate and the putamen. The putamen is anatomically and functionally connected with *the globus pallidus*, forming *the lentiform nucleus*, a lens-shaped structure separated into the putamen and the *external* and *internal globus pallidus*. The globus pallidus is recurrently connected with *the subthalamic nucleus*. Finally, lying in the midbrain of the mesencephalon is *the substantia nigra*. It is composed of the *pars reticulata* which receives fibres from the basal ganglia and output to the thalamus, and the *pars compacta* which project back to the caudate and the putamen through the nigrostriatal pathway creating an internal feedback loop within the basal ganglia system. The neurotransmitter involved is dopamine.

The striatum is characterized by three sets of afferent connections: from the cerebral cortex, from the intralaminar thalamic nuclei and from the pars compacta of the substantia nigra. The largest contingent of afferents comes from the cerebral cortex. Almost all regions of the cortex send fibres to the striatum, but the caudate nucleus and the putamen receive projections from different cortical regions. The putamen is dominated by somatotopically organised inputs from the primary sensory area and primary motor area, while the caudate nucleus receives projections predominantly from association areas. Thalamic inputs associated with a certain region project onto the same striatal region as does the comparable functional area of cortex (e.g., regions related to reward and motivation project to the ventral striatum, regions related to cognition project to the dorsal and medial caudate nucleus, and regions related to motor control project to the dorsolateral striatum). The striatum projects, in a generally topographic manner, primarily to the pallidal complex (not specified in **Figure 1.9**; it includes the external and internal segment of the global pallidus and the ventral pallidum of the substantia innominata), the substantia nigra, and the ventral tegmental area. The two most abundant neuronal cell types in the striatum are its projection neurons and the striatal interneurons. The projection neurons receive an excitatory glutamatergic input from the cortex and thalamus and a dopaminergic input from the substantia nigra pars compacta, as well as from striatal interneurons. They contain either the D1 or D2 type of dopamine receptor, responsible for the different effect of dopamine input to these cells. The D1 receptor

containing cells respond with excitation while the cells containing D2 receptors respond inhibitory to dopamine input (Cooper, Bloom, & Roth, 1996).

1.5.5.2 Corticostriatal circuits

The two main corticostriatal circuits (see **Figure 1.9**) are *the direct pathway* and *the indirect pathway*. The direct pathway passes through the cortex-striatum-GPi/SNr-thalamus-cortex and information that carried through this pathway has a net effect of excitation generating a positive feedback on the cortex. The indirect pathway connects similar to the direct pathway, but involves in addition the GPe and the subthalamic nucleus. It passes through the cortex-striatum-GPe-subthalamic nucleus-GPi/SNr-thalamus-cortex and has a net effect of inhibition generating a negative feedback on the cortex. Thus, the direct and indirect pathways affect the cortex (via the thalamic neurons) in opposite ways. This model of circuits has been used extensively to model how the basal ganglia may function in motor behaviour.

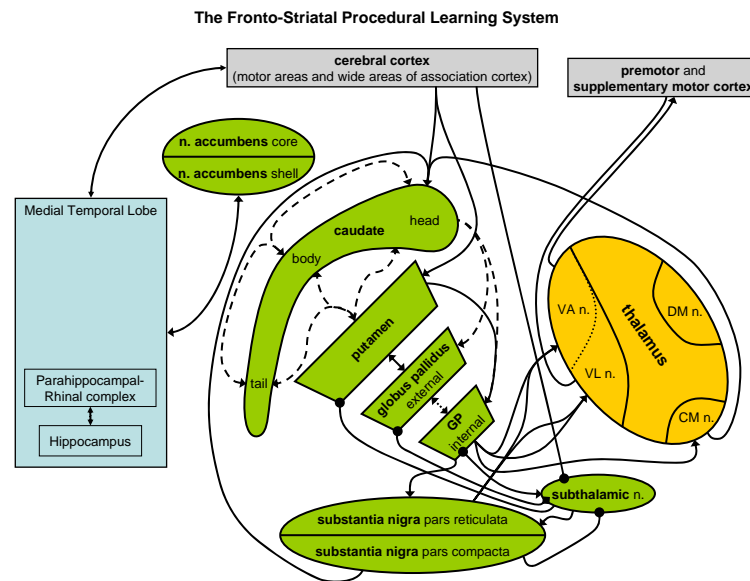


Figure 1.9 The main outline of the corticostriatal circuits that are thought to be involved in frontostriatal procedural learning.

The current idea of the functional and connectivity structure of the corticostriatal circuits suggests a segregated and parallel circuitry organisation. The basal ganglia, along with their connected cortical and thalamic areas, are viewed as components of parallel circuits whose functional and morphological segregation is relatively strictly maintained. Each circuit is thought to engage separate regions of the basal ganglia and thalamus, and the output of each appears to be centred on a different type of the frontal lobe. A well defined somatotopy is often maintained throughout the stages of the circuits, such that each circuit contains a number of highly specialized hierarchical flows of information that permit parallel, multilevel processing of a vast number of variables to process concurrently. The concept of parallel circuits help to explain the occurrence of the same symptoms observed caused by lesions at different locations in the cortico-striato-pallido-thalamo-cortical loop. In other words, one can relate to certain diseases as lesions of the frontostriatal circuitries. Behaviour syndromes observed with frontal lobe lesions can be related with striatal or thalamic lesions as recognizable circuit-specific behaviours.

According to the idea that the basal ganglia switch between actions, the basal ganglia perform an operation critical for shifting mental set. The ability to shift is required for both producing novel behaviour and for combining behaviour patterns into novel sequences of behaviour of higher complexity. In the motor domain, a problem in initiating movements can be viewed as a deficit in set shifting. The basal ganglia are in a position to monitor activation across wide regions of the cortex, allowing a shift between different actions and mental sets by removing an inhibitory influence in selected neurons. This shifting hypothesis also holds the key to the basal ganglia's role in learning. Dopamine is known to play a critical role in the reward system of the brain, providing the organism with neurochemical marker of the reinforcement contingencies that exist for different responses in the context of the current environment. Learning involves change in behaviour—either acquiring the appropriate response in an unfamiliar context or breaking a habitual response when contingencies change in a familiar context.

1.5.6 Neural correlates of sequence processing

1.5.6.1 Functional neuroimaging studies of natural and artificial syntax processing

The left inferior frontal region has been suggested to play a role in several cognitive domains, including for example working memory (Baddeley, 1986, 2003; Baddeley et al., 1998), language processing (Bookheimer, 2002; Friederici, 2002; Hagoort, 2003, 2005; Kaan &

Swaab, 2002), and musical syntax (Jackendoff, 2002; Lerdahl & Jackendoff, 1983; Patel, 2003). Additional evidence suggests that sequence learning depends on the left (Conway & Christiansen, 2001; Peigneux, Meulemans, Van der Linden, Salmon, & Petit, 1999) and the right inferior frontal region (Doyon, Owen, Petrides, Sziklas, & Evans, 1996). Moreover, it appears that the human left inferior frontal region is important for learning sequences which contain hierarchical structure (Gelfand & Bookheimer, 2003; Petersson et al., 2004). This suggests that the left inferior frontal region is engaged in the processing of structural aspects of cognitive representations and might provide a neural infrastructure for structural integration (Forkstam et al., 2006; Petersson et al., 2004). There is also evidence that the inferior frontal cortex is functionally subdivided (e.g., Bookheimer, 2002) which suggests some level of representational specificity as well as the possibility for interactive concurrent processing. Since none of these regions seem to be uniquely specific to linguistic syntax processing (see e.g. Kaan & Swaab, 2002) it is not unreasonable to suggest that syntax processing is in fact dependent on a functional network of interacting brain regions. This picture is similar to the one proposed by Hagoort (2003; 2005) in which the integration of various sources of linguistic information operates in parallel in a workspace for incremental unification of structured representation. Specifically, it is hypothesized that the left inferior frontal region subserves the unification space for language, one aspect of working memory.

With respect to language, the left inferior frontal gyrus and in particular the Brodmann's area (BA) 44 and 45 are implicated as a region involved in both syntactic comprehension and production (Bookheimer, 2002; Hagoort, 2003; Indefrey, Hagoort, Herzog, Seitz, & Brown, 2001). This notion finds support from both lesion data and functional neuroimaging investigations of natural syntax processing (cf., Stowe, Paans, Wijers, & Zwarts, 2004), as well as the distribution pattern of neural stimulations applied in order to disturb phonological, syntactic, and semantic processing (e.g., reviewed in Deacon, 1997). Syntax representations are likely created during comprehension even when the input is semantically un-interpretable. Two analyses of functional neuroimaging studies of language processing suggests that the most reliably replicable finding related to syntax parsing is localized to the left inferior frontal gyrus (BA 44 and 45) representing evidence for an involvement of Broca's region in aspects of syntax processing (Indefrey, 2004), and further, that there appears to be some evidence for a functional specialisation with respect to the left inferior frontal region. (Bookheimer, 2002). Even though there was considerable overlap, there seemed to be tendencies, at a coarse spatial scale, indicating that the anterior-inferior part of the left inferior frontal gyrus (centred around BA 47) is likely related to aspects of semantic processing, while the posterior-superior part (centred on the posterior parts of 44 and

extending into the anterior parts of BA 6) is likely related to aspects of phonological processing. The greatest spatial distribution was observed in the syntactically related activations. However, the centre-of-mass of this distribution appears to be localized to the middle part of the left inferior frontal gyrus centered on BA 44 and 45. However, the issue of precise spatial localisation in functional neuroimaging is complex and it appears that spatial precision in group studies of higher cognitive functions is on the order of approximately 10 mm (cf., Brett, Johnsrude, & Owen, 2002; Petersson, Nichols, Poline, & Holmes, 1999b).

The classical model for language organisation in the brain (Broca, 1861; Wernicke, 1874) relates language production to the anterior language regions in the dominant hemisphere, most commonly the left, centred on the posterior left inferior frontal region, and language comprehension to the posterior language regions centred on the posterior left superior temporal (restricted Wernicke's area) and surrounding parietotemporal regions (extended Wernicke's area; see **section 1.5.4**). However, this simple mapping of production and comprehension components onto anterior and posterior language related brain regions have since been re-examined (see e.g., Caplan, 1992; Kaan & Swaab, 2002; Zurif, 1990, 1998). Cortical electrical stimulation mapping has indicated that aspects of syntax processing may be related to the left posterior middle-inferior frontal, posterior superior temporal, and inferior parietal regions (Ojemann, 1983; Ojemann & Mateer, 1979). Also, several neuroimaging studies (for reviews see e.g., Bookheimer, 2002; Kaan & Swaab, 2002) have indicated that these regions may be associated with different aspects of syntax processing, including the syntactic complexity of sentences (Caplan, Alpert, & Waters, 1998, 1999; Caplan, Alpert, Waters, & Olivieri, 2000; Cooke, Zurif, DeVita, Alsop, Koenig, Detre, Gee, Pinango, Balogh, & Grossman, 2001; Inui, Otsu, Tanaka, Okada, Nishizawa, & Konishi, 1998; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Stromswold, Caplan, Alpert, & Rauch, 1996), grammatical error detection (Embick, Marantz, Miyashita, O'Neil, & Sakai, 2000; Indefrey et al., 2001; Kang, Constable, Gore, & Avrutin, 1999; Ni, Constable, Mencl, Pugh, Fulbright, Shaywitz, Shaywitz, Gore, & Shankweiler, 2000), or sentence matching under a syntactic/lexical manipulation (Dapretto & Bookheimer, 1999).

As noted earlier Chomsky has argued that children's capacity to acquire natural languages depends on an innate universal grammar (UG) that constrains the form of possible human languages (Chomsky, 1965, 1986; Chomsky & Lasnik, 1995). A study by Musso et al. (2003) attempted to investigate the neural correlate of acquiring new linguistic competence by teaching adult participants two types of rules, UG consistent rules and rules which have not been found in natural languages. They reported relative activation over time in Broca's region

for the former and relative decreased activation for the latter. These results are broadly consistent with the observations of Opitz & Friederici (2003). Musso et al. (2003) suggested that in Broca's region, biological constraints and language experience interact to enable new linguistic competence to develop. The subjects in the study of Musso et al. (2003) were explicitly taught the rules they had to learn, information was provided describing each rule with example sentences clarifying the rule. The subjects then practiced on correct and incorrect examples and performance feedback were provided. However, they were not provided with any information about phonological aspects of the new vocabulary. Thus, as noted by Marcus et al. (2003), one may ask whether the results reported reflect language acquisition as such. Marcus et al. (2003) also suggest a number of alternative interpretations in terms of working memory, complexity demands, or linguistically independent domain-general rule learning.

1.5.6.2 Implicit and explicit sequence learning and the medial temporal lobe

The medial temporal lobe memory system is critically involved in declarative and episodic memory (Cohen, Barch, Carter, & Servan-Schreiber, 1999; Eichenbaum & Cohen, 2001; Squire, 1992; Squire et al., 1993). The serial reaction time (SRT) task introduced in **section 1.3.9** involves several cortical and subcortical structures as investigated by fMRI and lesion studies (Daselaar, Rombouts, Veltman, Raaijmakers, & Jonker, 2003; Doyon, Penhune, & Ungerleider, 2003; Grafton, Hazeltine, & Ivry, 1995; Rauch, Whalen, Savage, Curran, Kendrick, Brown, Bush, Breiter, & Rosen, 1997), including motor regions, parietal regions, the basal ganglia, and the cerebellum, but implicit SRT has until recently appeared to be independent of the medial temporal lobe (MTL) memory system (including the hippocampus, cf. e.g., Curran, 1998). Other studies suggest however that the MTL may be involved in both implicit and explicit learning of visuomotor sequences (Schendan, Searl, Melrose, & Stern, 2003). The important commonalities in studies that report MTL activation is that they implement performance feedback to the subjects. For example, MTL activation have been reported during learning of material generated from artificial grammars/languages using a feedback approach (Fletcher, Buchel, Josephs, Friston, & Dolan, 1999; Opitz & Friederici, 2004; Strange, Henson, Friston, & Dolan, 2001). In a study of the weather prediction task (a category learning task in which good/bad weather is probabilistically determined by card-sequences), MTL activity was associated with receiving positive feedback but not with correct classification (Seeger & Cincotta, 2005), suggesting that the MTL might be involved in associative feedback prediction, perhaps based on sequence recognition at some level. Similarly, the explicit character of the task seems to be a likely explanation when observing learning related effects in the medial temporal lobe (Opitz & Friederici, 2003; Strange,

Fletcher, Henson, Friston, & Dolan, 1999). Results from a case study of a densely amnesic patient with bilateral MTL lesions (Gagnon, Foster, Turcotte, & Jongenelis, 2004) indicate that this region has a limited role in implicit learning of recurrent sequences, both in the SRT and Hebb's supra-span learning task (in which subjects echo digit strings in a series in which every third string is identical; learning is quantified in terms of increased accuracy on repeated strings compared to non-repeated). These findings are consistent with previous studies on implicit sequence learning (Reber & Squire, 1994; Reber & Squire, 1998) and artificial grammar learning (e.g., Knowlton & Squire, 1996; Knowlton & Squire, 1997) in amnesic patients as well as a rat model of the SRT task (Christie & Dalrymple-Alford, 2004). Similarly, using a computer version of the radial-arm maze (cf., DeCoteau & Kesner, 2000), Hopkins and colleagues (2004) found that amnesic subjects with selective MTL (hypoxic) lesions learned procedural sequences to the same degree as controls when compared to random sequences. In contrast, the controls performed significantly better than the amnesic patients on declarative sequences.

Whereas MTL lesions typically impair explicit but not implicit learning and memory, cases of implicit impairment and explicit sparing following basal ganglia lesions have been less consistent. It has been suggested that implicit and explicit sequence learning can proceed concurrently without interference (Jimenez & Mendez, 1999), although explicit processing can interfere with implicit learning (Schmidtke & Heuer, 1997), suggesting that explicit and implicit processes can interact during learning. As an example, in a study of standard and alternating SRT tasks, Fletcher and colleagues (2004) investigated the interaction between implicit and explicit learning. The results suggested right prefrontal, caudate nucleus, thalamus, and MTL engagement during sequence learning. The right prefrontal engagement seemed to be related to the explicit acquisition of alternating sequence structure. They suggested that explicit attempts to learn the alternating sequence reduced implicit learning and behaviour data indicated that the reduction in implicit acquisition was related to the suppression of learning itself rather than the expression of acquired knowledge. While they observed a negative frontothalamic interaction irrespective of task instruction with the standard SRT task, they observed a positive correlation between the right prefrontal region and the left thalamus in the explicit compared to the implicit alternating SRT condition.

1.5.6.3 Implicit and explicit sequence learning and subcortical structures

Implicit motor learning as indexed by the SRT task has been shown to be impaired in Parkinson patients (Smith & McDowall, 2004). However, in a study using dual-task conditions to reduce influence of attention and strategic learning, no difference between

Parkinson patients and controls was found (Kelly, Jahanshahi, & Dirnberger, 2004). Kim and colleagues (2004) found implicit SRT effect in controls but not in early stage Huntington patients, implicating the importance of intact basal ganglia structures in implicit motor sequence learning. In implicit artificial grammar learning no such deficit in early stage Huntington patients has been found (Forkstam, Voermans, Dekkers, Kremer, Fernández, & Petersson, In submission; Knowlton, Squire, Paulsen, Swerdlow, & Swenson, 1996b). In Forkstam et al. (In submission), despite similar behavioural performance, the Huntington patients showed an atypical fronto-striatal activation pattern compared to controls (see **study 4** of this thesis). In a study of individuals with unilateral cerebellar lesions (stroke) on an implicit motor learning task (sequence tracking) evidence for a cerebellar role in implicit motor learning was provided (Boyd & Winstein, 2004). They suggested that the cerebellum supports the formation of predictive strategies for the timing of motor responses and that this function is not lateralized but that cerebellar output may affect the formation of an internal model for timing movements in both extremities. Moreover, Torriero and colleagues (2004) showed that repetitive transcranial magnetic stimulation of the lateral cerebellum interferes with the acquisition of SRT task sequence knowledge and provided some evidence for hemispheric cerebellar differences with respect to the expression of learning. Representation of sequential structure can occur with respect to the order of perceptual events or the order in which actions are linked. Bischoff-Grethe and colleagues (2004) showed with event-related fMRI that transfer type (motor vs. perceptual) interacted with sequence retrieval (sequencing vs. rest) which revealed significantly greater activation in the bilateral supplementary and cingulate motor areas, ventral premotor cortex, left caudate, and inferior parietal lobule for subjects in the motor group suggesting successful sequence retrieval at the response level. Lungu and colleagues (2004) suggested that cortical brain regions including the prefrontal and motor regions encode the transitions from one element to the next early in learning, while the basal ganglia encode the full sequence structure toward the end of learning.

1.6 SUMMARY

Natural language acquisition is a largely spontaneous, non-supervised, and self-organised process. The structural aspects of natural language are acquired at an early age and largely without explicit feedback (Chomsky, 1965; Jackendoff, 2002; Pinker, 1994). In contrast, reading and writing are examples of typically explicitly taught cognitive skills (see e.g., Petersson, Ingvar, & Reis, 2009). The artificial grammar learning (AGL) paradigm has been proposed as a suitable model for the acquisition of structural aspects of language (Gomez & Gerken, 2000; Petersson et al., 2004) and for exploring differences between human and animal learning relevant to the faculty of language (Hauser et al., 2002). The underlying

grammar supports unbounded parsing and generation, and the paradigm comprise implicit learning on acquisition sets of grammatical examples alone without performance feedback (Forkstam et al., 2006; Petersson et al., 2004). It is likely that natural and artificial language acquisition share implicit acquisition mechanisms, as originally suggested by Reber (1967). Evidence from functional neuroimaging data is consistent with this suggestion. Brain regions related to natural language syntax are engaged in artificial syntax processing. In particular, the left inferior prefrontal cortex centered on Broca's region (Brodmann's area, BA, 44/45) is sensitive to artificial syntax violations (Forkstam et al., 2006; Petersson et al., 2004). This region is specifically sensitive to the structural properties rather than to local linear surface features of the input items. Additional support for the implicit character of artificial grammar learning comes from lesion studies on amnesic patients. Knowlton and Squire (1996) investigated artificial grammar learning in amnesic patients and normal controls on grammaticality classification. Both groups performed similarly on grammaticality classification, while the amnesic patients showed no explicit recollection of either whole-item or substring information, suggesting that artificial grammar learning depends on the implicit acquisition of structural knowledge (i.e., rule-based representations). Alternative theoretical frameworks have questioned the abstract (rule) acquisition picture and suggest instead that grammaticality classification utilizes exemplar-based representations (Vokey & Brooks, 1992) or substring representations (Perruchet & Pacteau, 1991). In order to address this issue and to control as well as test for any potential substring dependency, the ACS measure was developed (Knowlton & Squire, 1996; Meulemans & Van der Linden, 1997). Associative chunk strength (ACS) is a statistical measure of the associative familiarity of local substrings (e.g., bi- and trigrams) between a classification item and the acquisition set. It is quantified in terms of the frequency with which its substrings occur in the acquisition set. In this approach, acquired structural and instance specific information is quantified by grammaticality and ACS, respectively. From several studies which control ACS it is clear that structural knowledge is acquired (Forkstam et al., 2006; Meulemans & Van der Linden, 1997).

Evidence suggests that natural and artificial grammar learning can be conceptualized both in terms of structure based rule acquisition and surface based statistical learning mechanisms. We have proposed an alternative view on artificial grammar learning between these two conceptualisations (Forkstam et al., 2006; Petersson, Grenholm, & Forkstam, 2005) where proposal hierarchically structured information is recursively constructed from primitive structures which are stored in long-term memory. On-line integration of structured information could in this conceptualization result from the unification or successive merging of primitive structures, which are retrieved from long-term memory to the unification

component of working memory when activated. If a mechanism for on-line structural integration is already in place there is no need for a specific rule acquisition mechanism in order to establish a parsing process. Rule acquisition is instead accomplished by lexical acquisition of structured representations and their subsequent on-line unification (cf. e.g., Hagoort, 2004; Jackendoff, 2007; Vosse & Kempen, 2000). In this way the proposal re-traces a major trend in theoretical linguistics in which syntax is shifted into the mental lexicon in such a way that the distinction between lexical items and grammatical rules dissolves (Culicover & Jackendoff, 2005; Jackendoff, 2002, 2007).

The typical artificial grammar learning experiment includes a short acquisition session followed by a classification test. During the acquisition phase, participants are engaged in a short-term memory task using an acquisition sample of symbol sequences generated from an artificial grammar, typically a right-linear phrase structure grammar. Subsequent to the acquisition session, the subjects are informed that the items were generated according to a complex system of rules, without providing information about the rules, and the subjects have to classify new items based on their immediate intuitive impression (i.e., guessing based on gut-feeling), as grammatical or non-grammatical. The subjects typically perform reliably above chance, suggesting successful knowledge acquisition about relevant aspects of the underlying grammar. Subjects typically are unable to provide relevant motivations for their classification decisions (for reviews see Forkstam & Petersson, 2005; Seger, 1994; Stadler & Frensch, 1998), and this suggests that the classification performance is based on implicit acquisition mechanisms. However, it can be argued on principled grounds that invoking grammaticality judgments might not be the optimal way of accessing implicit knowledge, since the grammaticality classification instruction may induce a problem-solving strategy that, in principle at least, might encourage explicit processing (Manza & Bornstein, 1995; Newell & Bright, 2001). An alternative approach to probe implicit knowledge is based on the mere exposure effect. This effect refers to the finding that repeated exposure to a stimulus induces an increased preference for that stimulus compared to novel stimuli (Zajonc, 1968). For this reason in some of the studies of this thesis we use preference classifications to be able to keep the subjects completely unaware of the objectives of the task throughout the experiment.

CHAPTER 2 — METHODOLOGICAL BACKGROUND

2.1 WHOLE BRAIN INVESTIGATIONS

Cognitive brain function is a multi-level phenomenon and the study of brain function is consequently multi-disciplinary in nature. Detailed studies of adaptive processes and brain function will demand and benefit from a multi-disciplinary conceptual framework, using concepts from different scientific fields of research, including basic neuroscience, neuropsychology, cognitive neurosciences, linguistics, biological plausible and artificial neural network theory, machine learning and statistical learning theory (Arbib, 2003; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Gazzaniga, 1995; Gazzaniga, 2000, 2004; Rumelhart & McClelland, 1986; Schacter & Tulving, 1994a).

The methods chosen for the research described are functional neuroimaging methods that image brain functions (e.g., fMRI), in conjunction with behaviour methods from cognitive neuroscience and neuropsychology (e.g., classification and reaction time paradigms), as well as methods capturing the neurophysiological dynamics of human brain functions (e.g., electroencephalogram or EEG). Imaging methods are useful and well adapted for the study of dynamic processes in the human brain since it is possible to sample data with a reasonable resolution in time and space, a few seconds as well as a few millimetres respectively. This has the consequence that functional neuroimaging is a good tool to bridge the field between microscopic descriptions of the brain and its functions on one hand, and macroscopic descriptions of cognitive processes and behaviour on the other (Buckner & Koutstaal, 1998; Pascual-Leone, Bartres-Faz, & Keenan, 1999; Posner & Raichle, 1994).

Moreover, brain function is not only characterized by which regions are active but also by the interaction between different regions. With functional neuroimaging methods it is also possible to study the interactions between different brain regions. But to capture the hasty processes of networking neurons one has to combine functional neuroimaging methods with the more direct techniques such as EEG measurements of voltage fluctuations, or magnetoencephalogram measurements of magnetic field differences over time at the scalp (Menon, Ford, Lim, Glover, & Pfefferbaum, 1997; Opitz, Mecklinger, Friederici, & von Cramon, 1999).

Further inference for causal neuronal involvement in the processing of a cognitive function is derived from lesion studies (used e.g. in **study 4** of this thesis) and from studies on transcranial magnetic stimulations or TMS (Walsh & Pascual-Leone, 2003). However, in the domain of functional neuroimaging, neural regions correlating with dysfunction suggest a break in a functional system, not typically a suppression or modulation of a subfunction within that system. Similarly,

... "the removal of any of several widely spaced resistors may cause a radio set to emit howls, but it does not follow that howls are immediately associated with these resistors, or indeed that the causal relation is anything but the most indirect. In particular, we should not say that the function of the resistors in the normal circuit is to inhibit howling. Neurophysiologists, when faced with a comparable situation, have postulated 'suppressor regions' " (Gregory, 1961).

2.2 METHODS

In cognitive neuroscience, when investigating the living human brain, experimental strategies of cognitive and experimental psychology are combined with techniques allowing detailed investigations of brain activity correlating with cognition. In the process of constructing controlled experiments, our aim is that our results in the end will help to disentangle brain events that actually support cognition with brain events that merely correlate with unspecific (cognitive) processes. In this enterprise the problem is to capture the living brain with at the same time both time and space accuracy.

To succeed with such reversed engineering of brain function and brain structure we have to keep in mind the kind of correlational connectivity that underlie our methodological approach. We need to keep track on the scale of investigation, whether we investigate our objective best on the level of transmitter substance, receptor, membrane potential, neuronal spike or train of spikes, one-cell dendritic input and axonal output system, small group of neuronal network, or large-scale neuronal network. Finally, we need to have some idea what the effective behaviour variable might be, e.g. perception and sensory input categorisation, motor output adjustment, adjustment of reflexive activity, hormonal output control, emotional processes, or language processes.

In three of the studies of this thesis we used functional neuroimaging to correlate oxygen uptake of the neural substrate with the behaviour of the individuals. Functional neuroimaging methods provide experimental access to the living human brain. The techniques have

developed rapidly during the last few decades. A framework of well described theories and empirically validated methods of hemodynamically based functional neuroimaging methods are now available. Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are nowadays extensively used to investigate how neuronal processing correlates with changes in behaviour or cognitive processing (Frackowiak, Friston, Frith, Dolan, Price, Zeki, Ashburner, & Penny, 2004; Raichle, 1994). Any given method differ in its underlying assumptions and approximations employed, and the empirical data have to fulfil these assumptions and approximations for optimal use of the method. Of equal importance is the robustness of the methods used in analyzing the data. This notion emphasizes the importance of empirical validation and the explicit characterisation of the inherent limitations of a given method. Nevertheless, the standard functional neuroimaging methods provide useful means to investigate how networks of brain regions interact due to the whole brain coverage and the fact that primary data from different brain regions can be sampled in space and time on an approximately equal basis. When the underlying assumptions and limitations are taken into account, the various standard approaches used generally serve their purposes well (Pettersson, Nichols, Poline, & Holmes, 1999a; Pettersson et al., 1999b).

2.3 THE COUPLING BETWEEN NEURAL ACTIVITY AND REGIONAL CEREBRAL BLOOD FLOW

The neurophysiological basis of functional neuroimaging is the relatively tight and roughly linear coupling between the regional cerebral blood flow, the metabolic activity, and the neural electrophysiological activity (Gusnard, Raichle, & Raichle, 2001; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Rees, Friston, & Koch, 2000; Scannell & Young, 1999). At rest the human brain consumes approximately 20% of the oxygen and metabolic supply needed by the body, although the brain accounts for only approximately 2% of the body mass (Raichle, 2001). The oxygen is used in the oxidative metabolism of glucose to supply the brain with energy in the long-term (Raichle, 1997). Brief increases in neural activity of a given brain region implies that the energy and oxygen requirements in the given region increases and is accompanied by an increase in blood flow as well as glucose consumption that exceed the increase in oxygen consumption (Fox, Raichle, Mintun, & Dence, 1988). The relationship between oxygen consumption and blood flow is not proportional. In a region of momentary activity, the increase in glucose is partly broken down anaerobically by glycolysis despite of overcompensation in blood supply. As a result there is a lowered extraction fraction of oxygen that results in increased oxygen content in the blood nearby (Raichle, 2001). The study of Logothetis and colleagues (Logothetis, Guggenberger,

Peled, & Pauls, 1999; Logothetis et al., 2001) suggests that a spatially restricted increase in the blood-oxygen-level dependent (BOLD) FMRI signal directly reflects an increase in neural activity. When they recorded action potentials, local field potentials, and the BOLD signal in parallel, local field potentials correlated somewhat better with the observed BOLD signal than the action potentials. Local field potentials arise from the input to as well as integrative processes within neurons. These findings are consistent with autoradiographic measurements of glucose consumption by different brain regions in rats (Raichle, 2001). The signal-to-noise ratios for neural signals recorded directly from the brain are much greater than the accompanying FMRI signal (Raichle, 2001), suggesting that the absence of an FMRI signal does not necessarily imply an absence of information processing in a particular brain region.

The dominant energy consumption in the brain stems from maintenance at appropriate levels of ionic concentrations and membrane potentials. Neuronal signalling evokes ionic fluxes across membranes that need to be restored and most of these fluxes are supported directly or indirectly by the Na/K-ATPase and other ionic pumps (Siesjö, 1978). In FMRI, the sensitivity of the measured signal is based on the related increase in oxygenated blood locally. The BOLD signal correlates well with local field potentials generated to a large extent in the postsynaptic dendritic component where the restoration of large ionic fluxes is generated from the neuronal input. The signal is a composite net-activity and it is likely that regional activity is a mixture of both excitatory and inhibitory components. However, given that inhibitory signals give rise to hyperpolarization and less ionic leakage post-synaptically, the recorded signal might be more closely related to local excitatory activity (Shinohara, Dollinger, Brown, Rapoport, & Sokoloff, 1979).

2.4 IMAGE PROCESSING AND STATISTICAL ANALYSIS

The BOLD-FMRI data were subjected to several steps of image- and statistical analysis (**Figure 2.1**). Most functional neuroimaging studies are analyzed as group studies because we are typically interested in commonalities over participants that generalize to the population sample. Another reason for conducting group investigations is to increase the signal-to-noise ratio and thereby the statistical power of a given experiment. In all studies of this thesis the FMRI data were analyzed through out using the statistical parametric mapping software, SPM (<http://www.fil.ion.ucl.ac.uk/spm>). Functional neuroimaging data have a natural 3D structure and are pre-processed in two main steps. Data are first spatially normalised to account for inter-individual anatomical variation using spatial normalisation procedures. Data are then low-pass spatial filtered using 3D isotropic Gaussian convolution kernels to account

for residual anatomical variation and to increase signal to noise ratios. The typical way to analyse this type of data is a mass-univariate approach to each voxel (the lowest sample resolution volume) followed by subtraction analyses between conditions (Petersson et al., 1999a, 1999b).

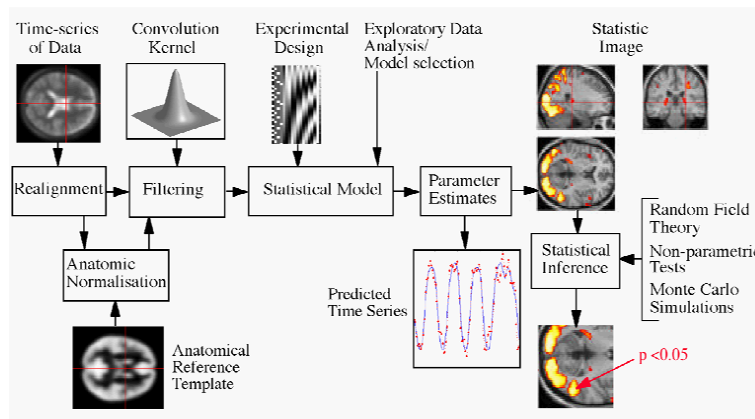


Figure 2.1 The general data analysis flow-chart in functional neuroimaging. The primary functional neuroimaging data are pre-processed (realigned, anatomically normalised, and spatially low-pass filtered), a statistical model for the data is created, model parameters are estimated and a test statistic is chosen in order to conduct statistical inference taking into account the multiple non-independent comparisons.

During the last decades a body of well described theories and empirically validated methods have been developed, providing a framework for investigating functional neuroimaging data and making scientific inferences based on statistical analysis. Statistical models make explicit as well as implicit assumptions about data. Of importance are how well these assumptions or approximations are fulfilled by the empirical data, and the robustness of the methods used when these are not fully met. The primary functional neuroimaging data are pre-processed (realigned, anatomically normalised, and spatially low-pass filtered), a statistical model for the data is created, model parameters are estimated and a test statistic is chosen in order to conduct statistical inference taking into account the multiple non-independent comparisons and possible auto-correlation (**Figure 2.1**). A short outline of the different processing steps involved in analyzing functional neuroimaging data will now follow.

2.4.1 Image pre-processing

2.4.1.1 *Realignment and anatomical normalisation*

In a functional neuroimaging study several measurements of 4d images (i.e., a time-series of 3d images) are acquired from several participants with the different experimental conditions occurring in a balanced order. In this thesis some hundreds BOLD-FMRI scans were acquired for each subject in each study. In order to minimize head movement during the experiment the participants were positioned in the MR scanner during the experiment with the head comfortably and softly fixated inside the scanner head coil. Small head movements on the order of 1 – 3 mm still occurred, and to compensate for this movement the images were realigned (cf., Ashburner & Friston, 1997) such that the brain in each volume of the time-series of a given individual occupied the same position in image space.

Since the brains of different individuals are anatomically different, a necessary requirement for group studies is to represent data in a standardized anatomical space. This requires anatomical normalisation to transform individual data into the standardized space. Anatomical normalisation aims to adjust for anatomical differences in order to allow data to be averaged across subjects and transforms the image time-series of the individual participant into a standardized anatomical space. In the studies of this thesis either the stereotactic space defined by the SPM template (<http://www.fil.ion.ucl.ac.uk/spm/>) or a study specific template was used.

2.4.1.2 *Functional-anatomical variability and spatial filtering*

There have been several attempts to assess the residual functional-anatomical variability after realignment and anatomical normalisation in more or less low-pass filtered data. These attempts have often used the variability in location of the local maximum statistic (peak location). Inter-subject standard deviations of the peak co-ordinates has been estimated to be on the order of 5-10 mm (Fox & Pardo, 1991; Hasnain, Fox, & Woldorff, 1998; Ramsey, Kirkby, Van Gelderen, Berman, Duyn, Frank, Mattay, Van Horn, Esposito, Moonen, & Weinberger, 1996). The intra-individual variability can also be significant, even for robust primary motor activations (Hunton, Miezin, Buckner, van Mier, Raichle, & Petersen, 1996). This variability increases when data from different laboratories are compared (Poline, Vandenberghe, Holmes, Friston, & Frackowiak, 1996; Senda, Ishii, Oda, Sadato, Kawashima, Sugiura, Kanno, Ardekani, Minoshima, & Tatsumi, 1998). This suggests that

activation foci that are less than 10 mm apart cannot be reliably distinguished (Grabowski, Frank, Brown, Damasio, Boles Ponto, Watkins, & Hichwa, 1996).

The inter-individual residual variability in functional anatomy generally exhibits spatial structure and is dependent on the algorithm used for normalisation. Simulation studies indicate that a reduction of registration (realignment) error and a minimization of the residual anatomic variability can significantly improve signal detection sensitivity (Worsley, Marrett, Neelin, Vandal, Friston, & Evans, 1996). In the presence of residual functional-anatomical variability the effect of inter-subject averaging amounts to a spatial filtering effect. Thus, if the spatial-scale of the filter matches the inherent scale of functional-anatomical variability in the population, no or little spatial information is lost. In general, using a voxel-based approach, it is important to reduce the impact of miss-registration and inter-individual residual functional-anatomical variability. A common strategy is to spatially low-pass filter the data either at reconstruction or with a suitably chosen convolution kernel (e.g., an isotropic 3D Gaussian kernel). Spatial filtering, which in effect is a local weighted averaging procedure, also increase the local equivalence of the voxel data across measurements and individuals and thus the validity of voxel-based statistical models.

Filtering data spatially may or may not increase the signal to noise ratio, depending on the relation between the size and shape of the signal and the convolution kernel used. This relation between the signal size/shape and the characteristics of convolution kernel can be understood in the light of the matched filter theorem (Rosenfeld & Kak, 1982). This theorem states that a signal in a background of white noise is detected with optimal sensitivity using a convolution kernel that matches the size and shape of the signal. It should be noted that the situation is slightly more complicated when the noise component is spatially autocorrelated, which typically is the case with hemodynamic functional neuroimaging data. The result of the matched filter theorem serves as a good approximation, if the spatial extent of the signal is large compared to the extent of the autocorrelation. If this is not the case the choice of an optimal filter is more complicated, and the autocorrelation has to be taken into account when choosing the filter. These issues apply similarly to signals in the temporal domain.

The objective of spatial filtering is mainly related to minimizing individual differences in residual functional-anatomy after anatomical normalization. Spatial filtering might also increases the signal-to-noise ration, since the power of the residual spatial noise is usually dominant in the higher spatial frequencies. Moreover, convolving the data with a Gaussian

kernel conditions the data to conform more closely to Gaussian random field theory. One of the mechanisms behind this is the fact that filtering amounts to a weighted averaging and it follows heuristically from the central limit theorem of probability theory that random variables averaged in this way converge towards a Gaussian distribution. In our studies, we have generally filtered the data with a 3D isotropic Gaussian kernel of 10 mm full-width-at-half-maximum (FWHM).

2.5 STATISTICAL MODELING AND ESTIMATION

2.5.1 The general linear model

Functional specialization or functional segregation is one central principle that has emerged in functional neuroimaging and in cognitive neuroscience at large. The idea of functional specialization rests on the hypothesis that different brain regions are specialized and implement different computations or operations on cognitive representations. This principle is reflected in the general linear model approach, a framework that encompasses all basic univariate models, including the ANOVA/ANCOVA and multiple regression models. In the general linear model framework n observations or response variables from a single image voxel are represented as the *data vector* Y of size $n \times 1$. The p effects and predictor variables are represented as p column vectors also of length n forming the *design matrix* X of size $n \times p$. The fixed *regression parameters* are represented as a column vector β of length p . The *residual random error* is written as the column vector ϵ of length n . It is typically assumed that ϵ is on the average zero and dependent (temporally autocorrelated and correlated as in repeated measures designs). The concise representation of the general linear model $Y = X\beta + \epsilon$ is:

$$E(Y) = X\beta, \text{Cov}[Y] = \text{Cov}[\epsilon] = T^2V,$$

where V is the $n \times n$ positive definite matrix. The usual normality assumption is only needed for statistical inference and no specific distributional assumptions are made. According to the Gauss-Markov theorem, the linear unbiased estimate of β that is best in terms of minimizing the squared estimation error is given by (least square estimation after pre-whitening):

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \text{ and } s^2 = 1/(n-p)(Y - X\hat{\beta})^T (Y - X\hat{\beta}),$$

$$\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y},$$

where $\hat{\beta}$ and s^2 are the estimate of the unknown β and σ^2 , respectively. Note that $\mathbf{Y} - \mathbf{X}\hat{\beta}$ is the residuals, and the form of s^2 is just the mean squared residuals (the $n-p$ reflecting the dimensionality of the residuals that are left after fitting p independent effects). Tests of linear combinations of the parameters can be made under the normality assumption, which gives:

$$\mathbf{C} \hat{\beta} \sim \mathbf{N}(\mathbf{C} \hat{\beta}, \mathbf{C}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{C}^T),$$

where \mathbf{C} is a row vector of length p , typically referred to as a **contrast** (cf., Frackowiak et al., 2004).

2.5.2 Baseline fluctuations and global normalisation

Functional neuroimaging experiments test hypotheses regarding regionally specific changes in neuronal activity. In BOLD fMRI these changes are indirectly reflected in the associated changes in regional susceptibility. Imaging experiments typically focus on relative regional differences. This can be problematic since variability in global factors often induces baseline fluctuations of both physiological (e.g., changes in $p\text{CO}_2$ levels and circulatory system changes) and of imaging system related origin (e.g., between-run variability in fMRI gain, see e.g., Frackowiak et al., 2004). Baseline fluctuations may be large, potentially hiding the effects of interest, and it is necessary to account for this variability in some manner. Measurements of global effects are predicated on the assumption that the variability in global effects adequately represent the baseline fluctuations and that the experimentally induced regional changes are superimposed on this. Several approaches to account for global changes have been developed, and how to explicitly model or remove the variability in baseline activity have been proposed and compared. It has been shown that most functional neuroimaging studies on normal subjects are roughly independent of the global normalisation method chosen (Aguirre, Zarahn, & D'Esposito, 1998). In this thesis we consistently used the proportional scaling approach. The problem of estimating baseline fluctuations should be less complicated if closely matched activation and reference conditions are investigated. The problem might become significant with increasing activation differences between conditions, emphasizing the need for carefully designed experiments that include active reference conditions.

2.5.3 Choosing reference state

If the activation and reference conditions differ in some relevant specific aspect of cognitive processing, the locations of statistically significant differences in signal between conditions will define brain regions related to this difference. The functional map generated by an activation condition will vary with the choice of reference condition, an adequate choice of comparative conditions is critical. In the simplest comparative analysis, the subtraction analysis, only parts of the underlying functional network will be observed while common components activated to a similar degree will not. Results obtained with the subtraction approach can only be interpreted as relative differences since a canonical reference state or baseline condition are difficult to define. This introduces a complication in the interpretation of functional maps, a fundamental ambiguity in the activation approach. Imagine a relative increase in regional cerebrospinal fluid in a certain condition A compared to another condition B. In relation to a third condition C this can either represent activation in condition A, a deactivation in condition B, or a combination of both.

The formulation and specification of the appropriate choice of reference condition(s) is an important and difficult issue and must be addressed at the design stage of the experiment. For a certain activation condition, the appropriate reference condition(s) depends crucially on the questions that the experimental data are supposed to address. With several reference conditions, multiple perspectives on the activation condition can be taken, for instance by using both closely matched control conditions and so-called low-level control conditions. With the closely matched control conditions that ideally differ only in a single aspect from the activation condition one can test for specifically induced effects, while with the low-level control condition (e.g., rest with eyes closed or visual fixation) one can detect common brain regions involved in different conditions.

The main complication in the interpretation of subtraction analysis results is under which comparisons a brain function is well approximated as a linear mixture of component effects. The interpretation of results from a subtraction approach is difficult and may strongly depend on the choice of experimental component tasks. This is particularly problematic if there is no canonical way of decomposing an overall task into components. It is necessary to develop new approaches that explicitly address the fact that brain functions emerge from non-linear interactions between components. Steps in this direction are the network analysis approaches of structural equations modelling (McIntosh & Gonzalez-Lima, 1994) and dynamical causal modelling (Friston, Harrison, & Penny, 2003). With the activation approach it is only possible to detect quantitative differences in information processing, or qualitative

differences accompanied by quantitative changes, network approaches can in principle detect qualitative differences in the pattern of interactions between brain regions without any changes in mean activity.

2.6 HYPOTHESIS TESTING AND STATISTICAL INFERENCE

The parameters in the general linear model are always assessed relative to their uncertainty in a statistical hypothesis-testing framework. What is searched for is if the magnitude of the contrast of parameter(s) is substantial with respect to its uncertainty, its standard deviation. In hypothesis testing the null hypothesis is assessed with a certain test statistic, a function of the data that is sensitive to departures from the null hypothesis and reflects the effects of interest. The observed statistic is compared to its distribution under the null hypothesis, yielding a P-value. A small P-value indicates that there is little support for the null hypothesis. The interpretation is more subtle however. The P-value is the probability of observing an equal or larger statistic value under an identical replication of the experiment assuming that the null hypothesis is true. It therefore is a statement about the data under the null hypothesis, neither the null hypothesis itself nor the alternative hypothesis.

In the decision theoretic framework for hypothesis testing, a pre-specified level of significance is used to accept or reject the null hypothesis (Bickel & Docksum, 1977). In alternative frameworks, the smallness of the P-value is viewed as a measure of the strength of the empirical evidence against the null hypothesis (Edgington, 1995). This perspective views the size of P-value as representing a smooth transition from empirical evidence supporting the alternative hypothesis to empirical evidence in favour of the null hypothesis. If one rejects the veracity of the null hypothesis whenever the P-value is below a critical value α then a valid test will control the false positive rate at α . The false negative rate β is on the other hand related to the statistical power, $1-\beta$, the probability of rejecting the null hypothesis when it is false.

The regression approach in functional neuroimaging fits univariate models at every voxel (the number of voxels is typically on the order of 10^5). Effects of interest are tested in each individual model by generating and assessing a statistic image, usually by fitting the same univariate model at each voxel. The test procedures conform to the standard structure of hypothesis testing, that is if a particular, pre-specified voxel is of interest, then standard univariate theory can be applied. Less regionally specific approach is to search the statistic image for voxels of significant magnitude using the local maximum statistic, or to search for

the significant clusters using the supra-threshold cluster size statistic given an intensity threshold.

The statistical analysis of functional neuroimaging data typically implies that many hypotheses are tested on the same data set. Central to this massive univariate hypothesis testing is an adequate handling of the multiple comparisons problem: It is necessary to appropriately control the false positive rate. The statistical inference procedure should handle the multiple comparisons problem effectively, avoiding any unnecessary loss of sensitivity and statistical power. Given the null-hypothesis and a test statistic of the data, the test is said to be liberal, conservative, or exact, if for any given threshold level and rejection region, the probability that the test statistics belongs to the rejection region is greater than, less than, or equal to the threshold level, respectively. Appropriate control of the false positive rate requires an exact or conservative test. However, the more conservative the test is, the lower is the sensitivity of the test. In order to handle the multiple comparisons problem (Hochberg & Tamhane, 1987) appropriately, the rejection criteria has to be chosen so that the probability of rejecting one or more of the null hypotheses when the rejected null hypotheses are actually true, is sufficiently small.

A common approach to control for multiple comparisons is the family-wise error rate, defined as the probability of falsely rejecting any of the null hypotheses. In the family-wise error approach evidence against the omnibus null hypothesis indicates the presence of some activation somewhere, meaning that the test has no localizing power and the false positive rate is not controlled for at individual voxels. Omnibus tests are tests with weak control of the family-wise error, useful to detect whether there is any experimentally induced effect at all, regardless of location. Instead a test procedure with strong control over the family-wise error is required to reliably locate the effect. This requires that the family-wise error is controlled not just under the null hypothesis but under any subset of hypotheses. Localizing power is ensured by a procedure testing all possible subsets of hypotheses with weak control over the family-wise error, making the test valid at every voxel, and the validity of the test in any given region not affected by the truth of the null hypothesis elsewhere.

Functional neuroimaging data are often characterized by spatial autocorrelation, meaning that closely spaced voxels are correlated, due to the imaging system, physiological factors, and post-acquisition smoothing of the image. Given a non-trivial spatial autocorrelation in the statistic image this implies that multiple comparisons are non-independent. An effective

solution of the multiple non-independent comparisons problem is central to the massive univariate approach. Broadly speaking, the approaches to handle this problem divide into parametric, non-parametric, and Monte-Carlo simulation approaches (for a review, see Petersson et al., 1999b). The parametric approaches used in functional neuroimaging are usually based on some type of random field theory (Adler, 1981; Worsley et al., 1996) which generate explicit distributional approximations.

The hypothesis testing and statistical inference in the studies of this thesis are based on parametric approaches founded in smooth random field theory. The random field theory approach allows for spatial correlation between voxels in the statistic image when correcting for multiple comparisons, preserving statistical power. It has proved versatile in testing a number of test statistics (e.g., local maximum, cluster size statistic, or the number of regions with size greater than a given size), and it has been extensively validated on simulated data. Empirical studies using real null data have indicated that this approach gives accurate results (e.g., Aguirre, Zarahn, & D'Esposito, 1997; Zarahn, Aguirre, & D'Esposito, 1997), and investigations of the robustness and characterization of inherent limitations of the approach with respect to the various assumptions and parameters have been carried out extensively.

In the application of smooth random field theory to discrete statistic images it is assumed that the statistic image can be considered as a well-sampled version of the smooth random field. In theory the spatial frequency spectrum of smooth stochastic process is not bounded, but in practice the observable spatial frequencies are limited in experimental data since only the spatial frequencies below half the frequency of the sampling process are observable by the Nyquist-Shannon sampling theorem. The sampling issue becomes particularly important in the context of smoothness estimation, which amounts to the estimation of a parameter related to the spatial auto-correlation. The estimation of the smoothness parameter should be independent of experimentally induced effects, which is why smoothness estimation is generally made on the residual images. In smooth random field theory, the volume, surface area, and diameter of an excursive set are estimated in the space of resolution elements (i.e., resel space, Worsley et al., 1996).

2.7 VOXEL BASED MORPHOMETRY

Voxel based morphometry is a neuroimaging analysis technique that allows investigation of focal differences in brain anatomy using the ordinary statistical approach describe above. In traditional morphometry, volume of the whole brain or its subparts is measured by drawing

regions of interests on images from brain scanning and calculating the volume enclosed. This is time consuming and can only provide measures of rather large regions while smaller differences in volume may be overlooked. In voxel based morphometry the grey and white matter segments of the structural MRI images are investigated (Ashburner & Friston, 2000, 2004, 2005; Good, Scahill, Fox, Ashburner, Friston, Chan, Crum, Rossor, & Frackowiak, 2002), characterizing local differences in the tissue density. Typical application of voxel based morphometry are group comparisons, or in longitudinal studies, on the same subjects. In any case, one has to take into account the importance of the impact of the image template. The typical hypothesis is that there is a difference between groups in specific regions. Creating a study specific template from all subjects included in the study minimizes the risk that the brains in one of the groups fit the template better. In this procedure, structural images from each brain are normalised to a common template to produce study specific grey and white matter templates by for example averaging over all normalised images. With this study specific template the segmentation procedure is performed anew, correcting the grey and white matter intensity levels using the Jacobian modulation. The images are then spatially filtered and statistical hypothesis testing is performed in a similar manner as in ordinary functional neuroimaging analysis.

CHAPTER 3 — SPECIFIC BACKGROUND TO THE INCLUDED STUDIES

3.1 STIMULUS MATERIAL

As stimulus material in all four studies of this thesis we generated grammatical (G) strings from the Reber grammar from the alphabet {M, S, V, R, X} (see **Figure 1.1**). In study 1, of the 110 possible grammatical items in the Reber language that can be generated of 2–8 letters, we randomly chose and allocated 56 items to the acquisition/training set and the remaining 54 items were included in the classification set. Non-grammatical strings were generated from the grammatical strings by randomly re-arranging the order of letters to render them non-grammatical. In this way we generated one set of 56 acquisition strings and one set of 108 classification strings (50% grammatical and 50% non-grammatical items). In study 2-4, we generated all possible strings with a string length of 5-12 letters length from the Reber grammar. Thus this string set was somewhat larger than in study 1, in total 569 strings. In addition to the grammatical/non-grammatical manipulation in study 1, we independently manipulated the expected familiarity of the surface structure of the strings by controlling for a familiarity score of 2 and 3 letter fragments in the classification strings in relation to the acquisition set. For each item we calculated frequency distribution of 2 and 3 letter chunks for both the terminal positions of the strings, as well as collapsed over all positions in the string, in order to derive the associative chunk strength (ACS) for each item (cf., Knowlton & Squire, 1996; Knowlton & Squire, 1997; Meulemans & Van der Linden, 1997). Then iteratively, we randomly selected 100 strings, generating an acquisition set which were comparable in terms of 2 and 3 letter chunks to the complete string set. For the remaining grammatical strings non-grammatical strings was generated by a letter replacement in two inner positions (not first or last position). The non-grammatical string selected were the ones best matching the grammatical strings in terms of both terminal and complete string position ACS (i.e., collapsed over order information within strings). These grammatical and non-grammatical strings were further classified as high/low ACS in terms of their ACS status independent of grammatical status, such that high/low ACS refer to classification strings composed of common/uncommon bi- and trigrams in the acquisition set, respectively. The classification sets were then randomly selected in an iterative procedure such that for each given classification set,

1. the high ACS strings in the classification and acquisition sets did not differ significantly ($P \gg 0.05$) in terms of ACS;

2. the high and low ACS strings did not differ significantly ($P \gg 0.05$) in terms of ACS when compared with the high and low ACS strings in the other classification sets; and
3. the low ACS strings **did** differ significantly ($P < 0.05$) in terms of ACS when compared with both the acquisition set and the high ACS strings in each classification set.

In this way, we expanded the classification sets and organised the classification material in a 2×2 factorial design with the factors grammaticality (grammatical/non-grammatical) and ACS (high/low), and the classification sets included 25% from each category: high ACS grammatical (HG), low ACS grammatical (LG), high ACS non-grammatical (HNG), and low ACS non-grammatical (LNG).

3.2 EXPERIMENTAL PROCEDURE

During acquisition the acquisition strings were presented on a computer screen either in its complete length (study 1 and 4), or letter-by-letter (study 2 and 4). The string presentation order was randomized for each acquisition session. After the presentation of each string on a monitor, when the string had disappeared from screen, the participants were asked to immediately type the string from memory in a self paced manner. The subjects were not provided with any sort of performance feedback and only positive examples (i.e., grammatical strings) were presented during acquisition. The classification sets were balanced across subjects, days and groups, and the string presentation order was randomized for each test. In the case of a baseline test this was presented to the subjects as independent of the subsequent testing. The preference classification groups were unaware of the final grammaticality classification test until it occurred on the last day, while participants with grammaticality instructions were informed about the existence of an underlying complex set of rules in the acquisition material during the days of the experiment (except for study 1 which took place within one day). Each classification session was presented as a yes/no classification task by inviting the subjects to respond with one of two response buttons after the finalized presentation of each string. The subjects were emphasised to respond with accuracy without haste, yet they had to keep up with the pace of the experiment.

3.3 BEHAVIOUR DATA ANALYSIS

For the behaviour data analysis we analysed the percent correct scores, the endorsement rates (i.e., strings accepted as grammatical or preferable regardless of grammaticality), and the statistical signal detection scores as measured with d' -prime (Macmillan & Creelman, 1991). Throughout the studies we used mixed-effect multi-way repeated measures analysis of

variance (ANOVA) using the statistics package R (www.r-project.org). We modelled the main factors grammaticality (grammatical/non-grammatical), associative chunk strength (high/low ACS), and, when applicable, classification session (test occasions) as within subjects fixed-effects. The main factor group, when included in the design, was modelled as between subject fixed-effect and subjects as random-effects. An overall significance level of $P < 0.05$ was used for statistical inference, and explanatory investigations for significant effects were restricted to the reduced ANOVA contrasted over the appropriate factor levels.

CHAPTER 4 — OVERVIEW OF THE EXPERIMENTS

The general objectives with the studies of this thesis were to develop an information processing model for the cognitive functions supporting implicit procedural learning. As criteria of the paradigm, it should be well described on Marr's (1982) three levels of description: the functional-computational level, the procedural-algorithmic level, and the implementational-hardware level (see **section 1.3.7**). The artificial grammar learning paradigm investigated in the seminal study of Arthur Reber (1967) was considered the best candidate for this purpose. Reber (1967) had in a behaviour setup investigated participants on a grammaticality classification task after acquisition of well-formed consonant strings generated from an artificial regular grammar. In parallel to the paradigm development, the use of this paradigm as a lab model for language acquisition was validated empirically.

All the studies of this thesis share the common denominator to investigate whether brain regions related to language processing overlap with the brain regions activated by a grammaticality classification task after acquisition of a small artificial grammar. The initial fMRI finding in study 1, that artificial syntax violations activated Broca's region (Brodmann's areas 44 and 45) of the left inferior frontal gyrus, was validated in study 2 using an elongated acquisition period and a stimuli material that was controlled for ACS (see **section 3.1**). In study 3 we investigated behaviourally the use of forced-choice preference classification comparing this with grammaticality classification (see **section 1.2.1**). In study 4 we tested the involvement of the caudate head in artificial grammar learning, a region suggested by the results in study 2 to be part of a frontostriatal network activated during artificial syntax classification.

4.1 STUDY 1: ARTIFICIAL SYNTACTIC VIOLATIONS ACTIVATE BROCA'S REGION

In the first study of the thesis (Petersson et al., 2004) we investigated whether the brain regions activated by the grammaticality classification task overlapped with brain regions related to natural language processing. It was suggested in a meta-analysis of functional neuroimaging studies on language processing (Bookheimer, 2002) that natural syntax processing is related to the left inferior frontal gyrus (Brodmann's areas 44 and 45) or Broca's region. We therefore designed our first experiment after behaviour piloting as a more elaborated replication of the study by Reber (1967). We used an implicit acquisition paradigm in which the participants were exposed to positive examples only. The stimulus material was generated as described in **section 3.1**. During acquisition each of the 12 subjects

studied the acquisition strings in a short-term memory task. They were instructed to attend to each consonant string as it were presented, in its whole length, on a computer screen for 5 s, and then as the string disappeared, to recall the string and type it into the computer using the keyboard. The paradigm was self-paced and subjects were allowed to correct themselves. The acquisition set was presented three times and the acquisition phase lasted approximately 40 min. Following the acquisition phase, the subjects were informed that the previously studied strings followed a complex set of rules and subsequently performed grammaticality classification (see **section 1.2.1**) in the scanner while event-related fMRI data were acquired. As a baseline condition for the fMRI analysis, we included a sensorimotor classification control task, in which the subjects had to indicate whether the presented string consisted of only P:s or L:s (controlled for average string length). The subjects responded by pressing one of two buttons in both grammaticality and baseline classifications. The classification and baseline items were presented in random order on a screen for 3 s during which the subjects responded by pressing a keypad, followed by a fixation-cross for 4 s. The computer screen was displayed to the subject by a LCD-projector standing inside the MR scanner room, projecting onto a semi-transparent projection screen that the subject viewed comfortably through a binocular device mounted on the head-coil.

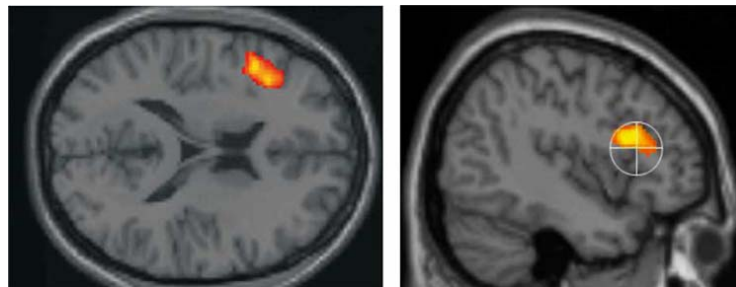


Figure 4.1 Artificial syntax violations were related to a significant activation in the left inferior frontal gyrus centred on Brodmann's area 45 extending into Brodmann's area 44 (Petersson et al., 2004).

The subjects showed significant above-chance correct classification performance on the classification task (mean \pm standard deviation = $73 \pm 7\%$, range = 61–92%, whereas 50% correct is expected by chance), validating that they were able to reliably differentiate between grammatical and non-grammatical items. We also observed that artificial syntax violations

activated Broca's region in all participants. More specifically, we observed that artificial syntax violations specifically activated Brodmann's areas 44 and 45 of the left inferior frontal gyrus (**Figure 4.1**). The violation effect was time specific in that the effect was stimulus-locked rather than response-locked (when time-locking on the subject responses we did not observe the effect).

In **Figure 4.1**, the cross-hair surrounding the region activated by artificial syntax violations is localized at the mean coordinates (approximately $[x\ y\ z] = [-44\ 19\ 12]$) of the natural syntax fMRI studies reported in the review of Bookheimer (2002). The mean distance of the individual local maxima reported in Bookheimer (2002) to the mean coordinates is approximately 13mm indicated by the radius of the circle. This suggests that the use of the knowledge acquired from an artificial grammar in an implicit acquisition paradigm, using only positive examples and no feedback, is subserved by the same neural processing infrastructure that has most consistently been related to human syntax processing. From this finding we concluded in study 1 that the activation of Broca's region in artificial syntax processing lends some support to the suggestions that artificial grammar learning provides a model for investigating aspects of structural language learning (Friederici et al., 2002; Gomez & Gerken, 2000), and exploring differences between human and animal learning relevant to the narrow faculty of language (Hauser et al., 2002). Further, that this finding validate the use of the artificial grammar learning paradigm to investigate syntax processing of hierarchical information structures in the human brain.

4.2 STUDY 2: NEURAL CORRELATES OF ARTIFICIAL SYNTACTIC STRUCTURE CLASSIFICATION

In the second study of the thesis (Forkstam et al., 2006) we modified and employed the artificial grammar learning paradigm to replicate and further investigate the role of the left inferior frontal cortex (Brodmann's area 44/45) and its sensitivity to artificial syntax violations. As in the first study, during each acquisition phase participants were engaged in a short-term memory task using an acquisition sample generated from the Reber machine (**Figure 1.1**). To make the presentation of the sequences more language like we changed to a sequential presentation setup, in which each string was presented symbol by symbol, instead of the parallel presentation setup used in study 1 where each string was presented in its whole. This had the further consequences that we also could measure EEG in parallel with the fMRI investigation. The sequential presentation has the further consequence that it increases the working memory load during processing of the stimulus material. This forces

the subjects' cognitive system to process the structured information sequentially, more similar to natural language processing.

Moreover, to investigate effect of extended acquisition and proficiency, the subjects participated in repeated acquisition sessions over 8 subsequent days. In addition, on day 1 and day 8 subsequent to the acquisition sessions, both event-related fMRI and EEG data were collected while the subjects performed grammaticality classification of novel strings. The objective with this manipulation was to correlate the behaviour performance with individual changes in functional neuroanatomy and event-related potentials (ERPs) over the learning period.

We also worked more intensively on the stimulus material as described in **section 3.1**. We expanded the classification sets, and in addition to the grammatical/non-grammatical manipulation we independently manipulated the expected familiarity of the surface structure of the strings. Thus, in study 2 we exposed the subjects to both a larger acquisition set and over a longer period than in study 1.

For the classification strings, we generated 4 sets of 56 (in total 112) grammatical and non-grammatical strings independently manipulated for high and low ACS. In this way we investigated the influence of substring familiarity in a factorial design, with grammaticality status (grammatical/non-grammatical) and associative chunk strength (ACS, high and low chunk similarity relative the acquisition strings, cf. Meulemans & Van der Linden, 1997) as factors. Thus the stimulus material included an acquisition set and four classification sets (all sets were pair-wise disjoint), and each classification set consisted of 28 strings of each string-type: high ACS grammatical (HG), low ACS grammatical (LG), high ACS non-grammatical (HNG), and low ACS non-grammatical (LNG).

It had been argued that sensitivity to the level of ACS is a reflection of a statistical fragment-based learning mechanism while sensitivity to grammaticality status independent of ACS is related to a structure-based acquisition mechanism (Knowlton & Squire, 1996; Meulemans & Van der Linden, 1997). Due to its involvement of frequency information, sensitivity to ACS manipulation suggests involvement of declarative learning mechanisms and hence a dependence on medial temporal lobe structures. Sensitivity to grammaticality status independent of ACS instead suggests involvement of procedural learning mechanisms, presumably dependent on interactions between prefrontal regions and the basal ganglia (cf.

e.g., Ullman, 2004 for a similar argument in natural language processing). With the factorial design of study 2 we had the means to investigate these claims.

The complete experiment included 8 days with one acquisition session each day followed, on day 1 and 8, by a classification test during which event-related fMRI data were acquired, and a subsequent classification test during which EEG data were acquired (see **section 4.2.1**). During both acquisition and classification sessions, each string was centrally presented letter-by-letter on a computer screen (2.7-6.9s corresponding to 5-12 letters; 300 ms letter presentation, 300 ms inter letter interval). During acquisition all subjects were presented with the 100 acquisition strings, randomly ordered for each acquisition session. When the last letter in a string disappeared the subject was instructed to immediately reconstruct the string from memory by typing on a keyboard in a self-paced fashion, similar to the procedure in study 1. Also in this study the subjects were not provided with any sort of performance feedback and only positive examples (i.e., grammatical strings) were presented during acquisition. During classification on day 1 and 8, the participants were presented with novel letter strings, in a similar way as during acquisition, while either fMRI or EEG measurements were conducted.

Different from study 1, this study introduced an instruction manipulation in that the subjects already in beginning of the experiment on day 1 were informed about the existence of a complex rule system underlying the strings composition. This mean that the knowledge of the task during the acquisition sessions differed between study 2 and study 1, in that the subjects were informed about the grammaticality classification task before any acquisition took part (see **section 1.2.1**). To minimize eye movements in the EEG setup, two white horizontal bars were presented centrally on a black background throughout the scanning sessions. In the fMRI setup a similar sensorimotor decision control task was included as in study 1. In this low-level baseline condition task, strings of repeated letters P or L (matched in string length to the classification set) were randomly interspersed and presented in the same fashion as the classification strings and subjects had to respond by pressing a button. The stimuli were presented via an LCD-projector standing outside the scanner room, projecting the computer display onto a semi-transparent screen that the subject comfortably viewed through a mirror device mounted on the head-coil.

The analysis of the classification performance in terms of endorsement rate (i.e., number of strings accepted as grammatical independent of grammaticality or ACS status) showed that

the subjects were sensitive to both grammaticality and ACS. The participants showed a greater sensitivity to grammaticality on day 8 compared to day 1, while they showed no additional sensitivity to ACS. Thus, subjects classified the items reliably above chance on the first day and the classification performance improved with repeated acquisition sessions (Figure 4.2).

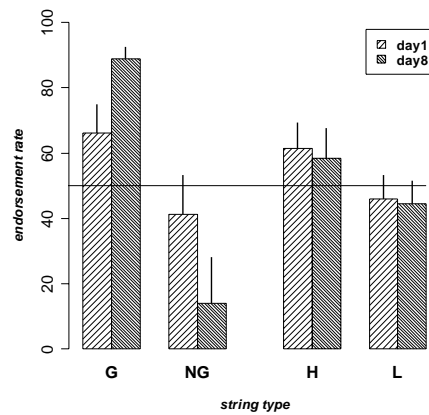


Figure 4.2 Endorsement rates (i.e., number of string classified as grammatical) over grammaticality and associative chunk strength (ACS). The endorsement rate for grammatical vs. non-grammatical items, but not for high vs. low ACS, increases as a function of repeated acquisition. (G = grammatical strings, NG = non-grammatical strings, H = high ACS strings, L = low ACS strings; error bars correspond to one standard deviation).

To be noted is that the proficiency in categorisation improved for each string category (i.e., HG, HNG, LG, LNG) over the 8 days. This shows that the subjects increased their detection capacity for both grammatical and non-grammatical strings independent of substring familiarity (Figure 4.3).

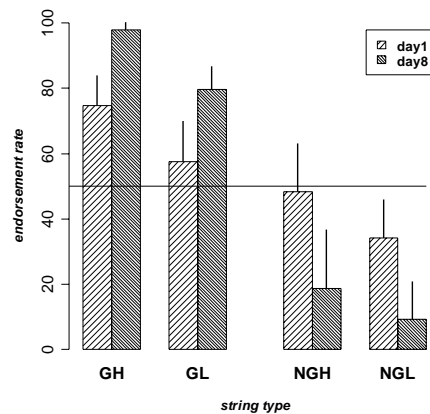


Figure 4.3 Endorsement rates (i.e., number of string classified as grammatical) over grammaticality and associative chunk strength (ACS) factor levels. The endorsement rate for grammatical vs. non-grammatical items increases as a function of repeated acquisition for both high and low ACS strings. (HG = grammatical high ACS strings, LG = grammatical low ACS strings, HNG = non-grammatical high ACS strings, LNG = non-grammatical low ACS strings; error bars correspond to one standard deviation).

In the second study, the fMRI data from 12 subjects performing reliably above chance on the grammaticality classification tasks on day 1 and 8 showed significant activation in a corticostriatal processing network compared to the sensorimotor decision baseline (**Figure 4.4**). These results were highly similar to study 1. Regions significantly sensitive to performance, grammaticality, substrings familiarity, or any interactions (omnibus ANOVA) included the inferior frontal region, the frontal operculum and anterior insula bilaterally, as well as the caudate nucleus bilaterally. To investigate which experimental manipulations carried explanatory value for the peak nodes in this network, we performed a regions-of-interest analysis on the cluster local maxima.

In this analysis we found the prefrontal regions activated more extensively than in study 1 and that the activated regions also included right homotopic regions. The Broca's region was specifically sensitive to artificial syntax violations, meaning that we replicated the findings from study 1 with high specificity. In particular, the left inferior frontal region (BA 44/45) was specifically related to syntax violations. It was the only frontal region which was

selectively sensitive to grammaticality but not to the level of associative chunk strength (ACS). This lends support for the suggestion that the left inferior frontal cortex (BA 45) has a specific role in processing structural regularities. This is also consistent with results showing that the left prefrontal cortex subserves syntax processing independent of lexical meaning (Indefrey et al., 2001). In contrast, the right inferior frontal region was sensitive to ACS, potentially reflecting aspects of declarative retrieval or generic error detection processes (cf., Indefrey et al., 2001). This finding was also the motivation behind a follow-up investigation on the role of the left and right inferior frontal regions classification of artificial syntactic structures by means of an off-line repetitive transcranial magnetic stimulation (rTMS) paradigm (Uddén, Folia, Forkstam, Ingvar, Fernandez, Overeem, Elswijk, Hagoort, & Petersson, 2008). In this study, the participants showed a reliable effect of grammaticality on classification of novel items after 5 days of exposure to grammatical exemplars without performance feedback in an implicit acquisition task. The stimulus material was manipulated independently for grammaticality status and local substring familiarity (see **section 3.1**). The results show that rTMS of BA 44/45 improves syntactic classification performance by increasing the rejection rate of non-grammatical items and by shortening reaction times of correct rejections specifically after left-sided stimulation, suggesting that activity in the inferior frontal region is causally related to artificial syntax processing.

In addition to the left inferior frontal activation, the head of the caudate nucleus correlated positively with syntactic correctness. This activation was however only present in the later part of the acquisition period, on day 8 but not on day 1. At the same time the subject behaviour performance was higher on day 8 relative day 1. This suggests a higher fluency for the subjects in the task, and the activation of the caudate head then suggests that this region contributes to an increase in cognitive processing fluency. This finding motivated the **study 4** of this thesis, using Huntington's disease as a neurological model for the involvement of the caudate head in artificial grammar learning.

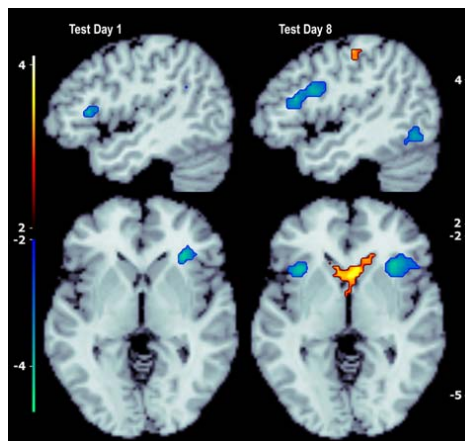


Figure 4.4 Regions sensitive to grammatical status (day 1 and 8; grammatical > non-grammatical items in red, non-grammatical > grammatical items in blue; only correct response trials included). Left: test day 1. Right: test day 8. Threshold at $P = 0.05$ corrected for multiple non-independent comparisons using the false discovery rate. Activations are projected onto the normalised structural (T1) image from a single subject ($x = -45; z = -3$; left is left; upper slices are in the left hemisphere).

4.2.1 ERP correlates of artificial syntactic structure classification

The second study also included an EEG experiment. Our aim with this approach was to also electrophysiologically relate the artificial grammar learning paradigm with natural language processing, to validate the usefulness of the paradigm as a tool to explore syntax processing in language. We set out to explore modulation of well-known language related ERP components and their time-frequency patterns during syntactic classification. A primary achievement in the understanding the functional neurophysiology of the human language faculty has been the revelation of neurophysiological processes associated with the treatment of semantic vs. syntactic aspects of language, through the analysis of ERPs during word and sentence processing. Over the last two decades, at least three specific language related ERP components have been identified and analyzed, including the N400, P600/SPS and the LAN. The N400 is an enhanced centroparietal negative-going component for semantically inappropriate words, peaking at about 400 ms after the offending word (Kutas & Hillyard, 1980). The P600 (Osterhout & Holcomb, 1992) or syntactic positive shift (SPS, Hagoort, Brown, & Groothusen, 1993) is a late centroparietal positivity associated with the processing of syntactic anomalies. The left anterior negativity (LAN) is a late negative-going potential

that can occur between 300 and 700 ms, with a left anterior spatial distribution, associated with syntactic structural processing complexity (Kluender & Kutas, 1993). Processing consistently shown to elicit a LAN effect includes the processing of *content words* that carry meaning (i.e., open-class words including nouns, verbs and adjectives) vs. *function words* that are essentially empty of meaning but contain information that is crucial for syntactic parsing (i.e., closed-class words including prepositions or determiners).

The N400 and SPS/P600 are traditionally associated with language processing, and both have also been evoked in appropriate non-linguistic cognitive sequence processing tasks. For the N400, it appears that the effect is evoked only if a particular stimulus does not fit into a pre-established semantic context independent of the input code (Niedeggen, Rosler, & Jost, 1999). Similarly, the SPS/P600 can be evoked in linguistic and non-linguistic conditions in which an element of a rule governed sequence is difficult to integrate, including the processing of musical phrase structure violations (Patel, Gibson, Ratner, Besson, & Holcomb, 1998), and in the processing of non-linguistic sequences that violate a learned rule in an artificial grammar task (Lelekov, Dominey, & Garcia-Larrea, 2000). We now wanted to validate and explore this finding in our artificial grammar learning paradigm with elongated acquisition.

We present here the analysis of the ERP data from 19 of the subjects included in study 2 performing reliably above chance on grammaticality classification on day 1 and 8. The EEG classification sessions occurred subsequent to the fMRI classification sessions, and the four classification sets was balanced over the two fMRI sessions and the two EEG sessions. Similar to the fMRI sessions each string was centrally presented letter-by-letter on a computer screen (2.7-6.9s corresponding to 5-12 letters; 300 ms letter presentation duration, 300 ms inter-letter-interval) using the Presentation software (<http://nbs.neuro-bs.com>). The acquisition sessions were the same as in the fMRI part of the study. During classification on day 1 and 8, the participants were presented with novel letter strings while EEG measurements were acquired using a 61-channel scalp electrode BrainVision amplifier setup. As in the fMRI sessions the subjects were informed about the grammaticality classification task before the first acquisition session. After a 1 s pre-stimulus period the strings were presented between these bars, followed by a 1 s delay and 2.5 s fixation cross period, indicating that the subject had to respond by pressing either left or right index finger, balanced within subject over sessions. The classification sets and string presentation order were balanced over subjects. The stimuli were presented via a computer monitor standing

inside a faraday cage room, comfortably viewed by the subject from approximately 1 metre distance.

In the ERP analysis we applied the cluster randomization algorithm implemented in the FieldTrip software (<http://fieldtrip.fcdonders.nl>) to keep the multiple comparison problem under control (Maris, 2004). Essentially two different approaches were made to disentangle the time signal in the EEG data as the subjects classified grammatical and non-grammatical items. In the first approach we performed a time-locked analysis, time-locked to the initial violation position in the letter strings. Recall that each non-grammatical string differed from a matched grammatical string by sequential violation in two locations in the string. The information up to the point of the initial violation was exactly the same in the grammatical and the non-grammatical string, which suggests that any deviation time-locked to this event likely has been triggered by the detection of a grammaticality violation. Indeed, strings with a syntax violation in the initial part of the string did show up as a late positive deviation during classifications on day 1 (**Figure 4.5**). However, the P600/SPS component was absent on day 8.

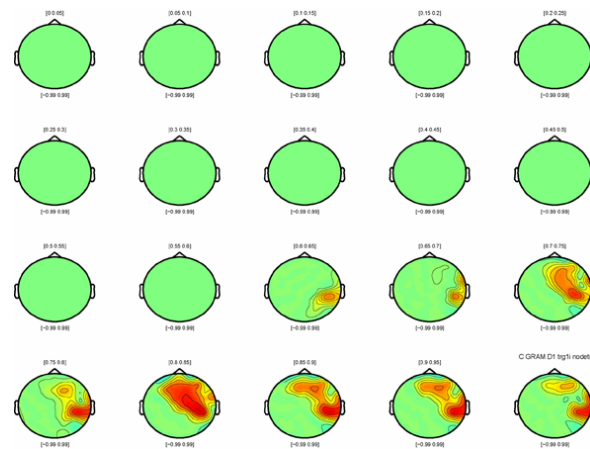


Figure 4.5 Day 1 topographic ERP distribution for non-grammatical strings compared to grammatical strings; time-locked to the first violation position; cluster-randomization analysis (two-sided t-test), baseline [-150,0] ms. The positive deviation starts at 500 ms after the presentation of the critical letter in the syntax violation target position.

In the second approach we performed a time-locked analysis, time-locked to the last letter in the sequence. In this analysis a negative deviation for grammatical violations was found on both day 1 (**Figure 4.6**) and day 8 (**Figure 4.7**). The effect was larger and lasted over a longer time period on day 8, as indicated by the significantly larger negative shift on day 8 as compared to day 1 in the later part of the component (**Figure 4.8**).

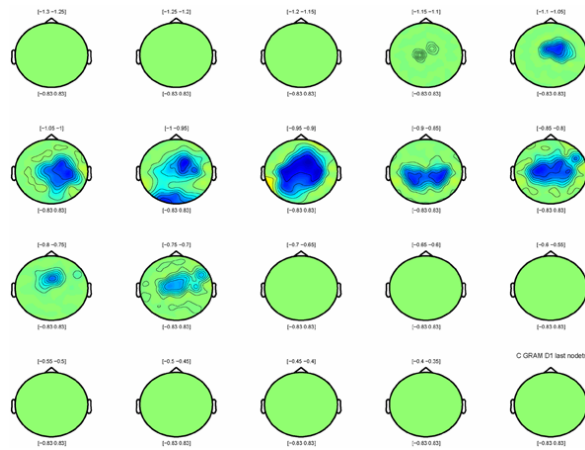


Figure 4.6 Day 1 topographic ERP distribution for non-grammatical strings as compared to grammatical strings, time-locked to the last letter position; cluster-randomization analysis (two-sided t-test), baseline [-150,0] ms. The negative deviation on day 1 lasts from 100 ms until 500 ms after the last letter is presented.

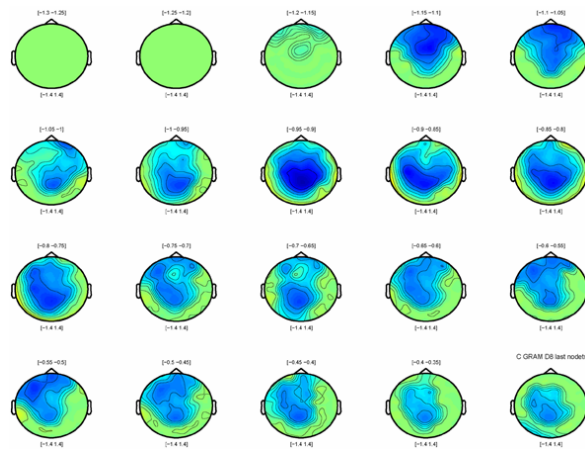


Figure 4.7 Day 8 topographic ERP distribution for non-grammatical strings as compared to grammatical strings, time-locked to the last letter position; cluster-randomization analysis (two-sided t-test), baseline [-150,0] ms. The negative deviation on day 1 starts already 50 ms after the last letter is presented.

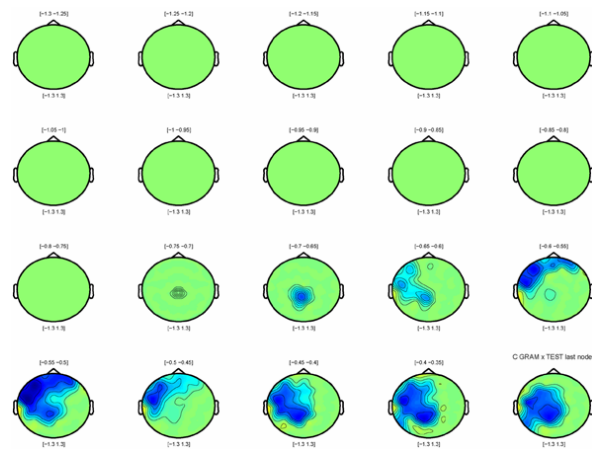


Figure 4.8 Topographic ERP distribution for the difference in ERP between day 1 and day 8 for non-grammatical strings as compared to grammatical strings; time-locked to the last letter position; cluster-randomization analysis (two-sided t-test), baseline [-150,0] ms. This figure illustrates the additional negative deviation on day 8 as compared to day 1 (see **Figure 4.6** and **Figure 4.7**).

The outcome of the EEG experiment suggests that it is possible to detect differences in the EEG signal related to the violation position. Strings with a syntax violation in the initial part of the string did show up as a late positive deviation during classifications on day 1 (**Figure 4.5**), lending some support to the previous artificial grammar learning finding (Lelekov et al., 2000) that artificial syntax violations trigger a late positive shift similar the P600/SPS component. However, the P600/SPS component was absent on day 8. This is unexpected if taking the view that skill acquisition, in this case regularity detection working on acquired knowledge of syntactic structure, should be reflected as an increase in relevant components to index the phenomenon. The possibility remains that the failure to replicate the P600/SPS was a false negative, possibly due to lack of statistical power. The population sample used in the

EEG investigation was however larger than the one used in the parallel fMRI investigation. A significant EEG difference was also observed time locked to the last presented letter on both day 8 and day 1, and reliably more so on day 8 than on day 1. This effect time-locked to the last letter presentation shares some similarities with the N400 component but it is unclear what cognitive processing this effect relates to. In the case the component is similar to the N400 this suggests that also non-syntax related components can get elicited in artificial grammar learning, and that this effect increases with performance. This interpretation is congruent with the suggestion that N400 gets elicited when a particular stimulus does not fit into a pre-established semantic context (Niedeggen et al., 1999). We are currently investigating and exploring the use of the repeated artificial grammar learning paradigm in follow up EEG studies with higher statistical power using both larger classification sets and more subjects.

4.3 STUDY 3: INSTRUCTION EFFECTS IN IMPLICIT ARTIFICIAL GRAMMAR LEARNING: A PREFERENCE FOR GRAMMATICALITY

In study 3 we investigated the behaviour effect in artificial grammar learning repeated over several days, while taking advantage of the structural mere exposure effect (Manza & Bornstein, 1995). In mere exposure artificial grammar learning subjects receive a preference classification instruction (see **section 1.2.1**) which make no reference to any previous acquisition episode and the subjects are not informed about the existence of an underlying generative mechanism. The idea is that mere exposure artificial grammar learning might measure implicit knowledge in a more pure manner because there is nothing in the classification procedures that refers to the acquisition part of the experiment and no reference to a complex set of rules are ever made. It has been shown that the preference classification instruction induces similar classification performance as the grammaticality instruction (Buchner, 1994; Manza & Bornstein, 1995) in a graded classification task (i.e., preference continuum). However, the graded classification complicates a direct comparison with forced-choice (yes-no) grammaticality classification.

The typical artificial grammar learning setup employs grammaticality classification instruction (see **section 1.2.1**). Some researchers have raised the concern that this might direct subjects to use explicit problem solving strategies, based on perceived regularities or imagined rules, and that this might interfere with the implicitly acquired knowledge put to use during classification. This concern appears unwarranted on theoretic grounds as long as incidental implicit acquisition is employed and the subjects base their classification decisions on their immediate ‘gut-feeling’. However, if performance on preference and grammaticality

classifications is statistically similar also empirical data would indicate that similar concerns are of less importance. The objectives of study 3 was therefore to compare forced-choice (yes-no) preference with grammaticality classification and to investigate whether and to what extent preference classification would show a similar pattern of results as the standard grammaticality classification. If no differences between forced-choice preference and grammaticality classification are found, then at least from an observational point of view grammaticality and preference classifications are similar as quantification measures for the acquired knowledge in artificial grammar learning. The use of preference or grammaticality classification could then be more a matter of which instruction that fits the experimental setup best.

Just as in study 2, the stimulus material was organised in a 2 x 2 factorial design using the factors grammaticality (grammatical/non-grammatical) and level of associative chunk strength (ACS; high/low; see **section 3.1**). Thus we are able to assess differences between instruction types related to grammaticality as well as of substring familiarity (i.e., ACS). In addition to the rate of acquisition, we also investigated the influence of instruction type on pre-acquisition baseline classification. To investigate the effect of instruction during baseline, we pooled the participants from experiments 1-3 (i.e., 20 participants with grammaticality classification and 20 with preference classification) and divided the baseline items into two equal sized time-blocks of 20 items (first/second half as they were presented over time). The basic ANOVA was extended with the factor block [1/2]. For the random acquisition set in experiment 2, 100 random strings were generated from the same alphabet and of the same length and with similar levels of associative chunk strength as the Reber grammar acquisition set.

In experiment 1 we directly compared the outcome of implicit artificial grammar learning in subjects given either the grammaticality classification or preference classification instruction. In this experiment subjects engaged in a short-term memory task using only grammatical strings without performance feedback for 5 days. Classification performance was independent of instruction type and both the preference and the grammaticality group acquired relevant knowledge of the underlying generative mechanism to a similar degree (**Figure 4.9**).

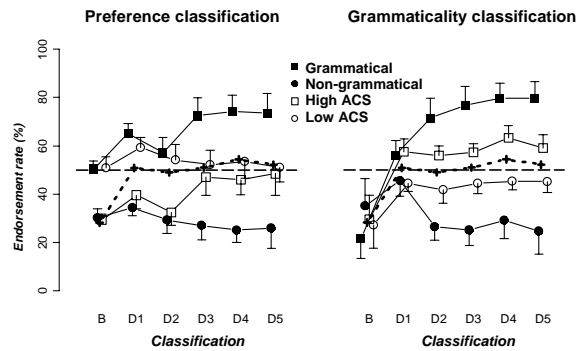


Figure 4.9 Experiment 1: Classification performance day 1-5. Endorsement rate (mean and standard error) as a function of instruction type; B = baseline, D1-5 = day 1-5 classification, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50 % chance level (straight dotted line).

In experiment 2, changing the grammatical strings to random strings in the acquisition material resulted in classification being driven by local substrings familiarity (**Figure 4.10**). In experiment 3, contrasting repeated vs. non-repeated preference classification showed that the effect of local substrings familiarity decreases with repeated classification (**Figure 4.11**). This was not the case in the repeated grammaticality classifications. We conclude in this study that classification performance is largely independent of instruction type and that forced-choice preference classification is equivalent to the typical grammaticality classification.

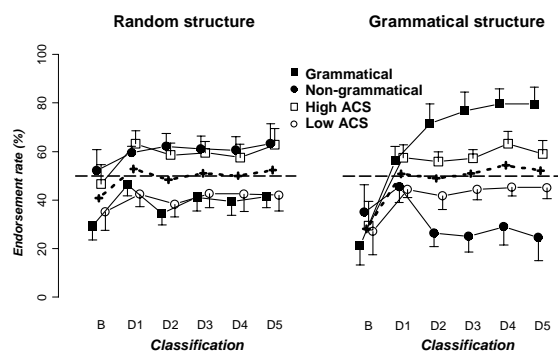


Figure 4.10 Experiment 2: The effect of acquisition on Reber and random strings. Endorsement rate (mean and standard error) as a function of acquisition material; B =

baseline, D1-5 = day 1-5 classification, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50 % chance level (straight dotted line).

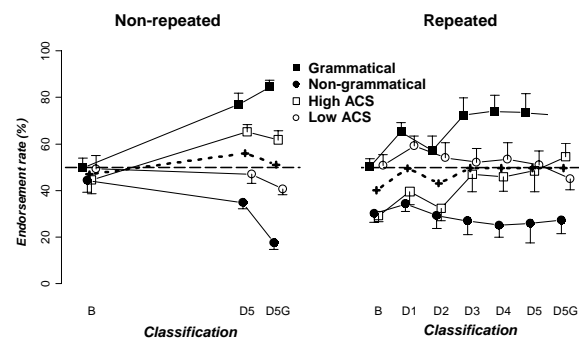


Figure 4.11 Experiment 3: The effect of repeated classification. Endorsement rate (mean and standard error) as a function of preference classification repetition; B = baseline, D1-5 = day 1-5 classification, D5G = final grammaticality classification, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50% chance level (straight dotted line).

The main finding of study 3 is that forced-choice preference classification is behaviourally equivalent to the standard grammaticality classification, suggesting that implicit acquisition of knowledge about the underlying generative mechanism are assessed similarly in the preference and the grammaticality version of artificial grammar learning. This is most clearly suggested in the non-repeated preference classification group. This group had a final grammaticality classification test on the last day and was never informed about the existence of a complex set of rules for generating the acquisition set until their final grammaticality classification test. This group showed the same overall pattern of results as the grammaticality group and the repeated preference classification group (recently replicated in e.g. Folia, Uddén, Forkstam, Ingvar, Hagoort, & Petersson, 2008). Thus, the artificial grammar learning paradigm yields very robust implicit learning quite independent of the experimental details. However, we suggest that forced-choice preference classification might have certain theoretical advantages over grammaticality classification. It appears that preference classification induces less dependency on surface features related to local

substrings. Preference classification might thus be less likely to induce explicit problem solving strategies, if these latter possibilities are of a real concern. In addition, the grammaticality effect was boosted in the non-repeated preference classification group when switching from preference to grammaticality instruction (a finding recently replicated in e.g. Folia et al., 2008). This might suggest that the grammaticality instruction is perceived as more well-defined by the subjects. Alternatively, the grammaticality instruction might trigger general motivation or attention effects. As a final remark, we note that the fact that effects of grammaticality as well as associative chunk strength can develop already during classification prior to acquisition suggests that the mechanism engaged (not necessarily the same as in artificial grammar learning proper) can work on surprisingly scarce input.

4.4 STUDY 4: FRONTOSTRIATAL CIRCUITRY IN ARTIFICIAL GRAMMAR LEARNING: AN FMRI STUDY IN HUNTINGTON'S DISEASE

Several studies suggest that artificial grammar learning involve a frontostriatal network comprising the head of the caudate nucleus and the left Brodmann's area 44/45 (e.g., Lieberman et al., 2004, see also study 2 of this thesis). In study 4 of this thesis we investigated the consequences of a deteriorating caudate infrastructure on successful artificial syntactic classification using event-related FMRI in early stage Huntington's disease (HD). The caudate nucleus is a core region affected in HD, a neuropathological well-characterized genetic neurodegenerative disease which in its early stages consists of a gradual, selective neuronal loss with concomitant gliosis of the caudate nucleus and the putamen, in combination with only little or moderate macroscopic atrophy (Vonsattel, Myers, Stevens, Ferrante, Bird, & Richardson, 1985). The relatively well-defined caudate degeneration in the early stage of the disease suggests that the role of the caudate nucleus in grammatical processing can be investigated with implicit artificial grammar learning, and raises the question how incipient functional caudate changes affect the network proposed to be functional in procedural memory. The use of artificial grammar learning to assess non-motor procedural memory may circumvent the limitations imposed by motor impairments related to HD. If the caudate nucleus plays a causal role in the neural architecture supporting grammatical processing and language acquisition, caudate degeneration in early stage HD implicate classification impairment in implicit artificial grammar learning. However, Huntington patients was previously shown to perform similar to controls (Knowlton et al., 1996b). Unfortunately Knowlton and colleagues confounded grammaticality with substring familiarity in such a manner that the normal classification performance could be explained by recognition of familiar substrings supported by medial temporal lobe declarative memory processes not typically affected in HD.

The objectives with study 4 was to clarify the caudate involvement in implicit artificial grammar learning using test material not confounded with substring familiarity (see **section 3.1**). Participants engaged in implicit acquisition of grammatical consonant strings over three days (full length string presentation; see **section 3.2**). On day 3 the subjects performed preference classification inside the scanner, followed by grammaticality classification outside the scanner (see **section 1.2.1**).

Both controls and Huntington patients performed similarly and reliably above chance on both tasks, extending earlier findings of intact artificial grammar learning in Huntington's disease to stimuli material controlled familiarity of substrings (Knowlton et al., 1996b). Despite similar behaviour performance the Huntington patients showed an atypical frontostriatal activation pattern compared to controls: basal ganglia structures outside the caudate head activated as a function of the experimental manipulations, hippocampus activated as a function of grammaticality and not only as a function of substring familiarity, and the left Brodmann's area 44/45 activated in the Huntington patients also as a function of substring familiarity and not only for grammaticality as in the controls. The hippocampus effects of grammaticality in the Huntington group but not in controls suggests that the medial temporal lobe memory system might compensate for a deteriorating caudate, consistent with the notion of a cooperative interaction between the medial temporal lobe memory system and the basal ganglia learning system. We suggest that the caudate deterioration in early-phase Huntington's disease reveal a cooperative and compensatory interaction between the medial temporal lobe memory system and the basal ganglia learning system, in order to keep performance levels during artificial grammar learning classification roughly intact (consistent with e.g. Voermans et al., 2004). We conclude that the results of the present study are consistent with the notion that the caudate is involved in artificial grammar learning classification.

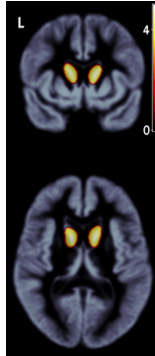


Figure 4.12 Voxel based morphometric grey matter reduction in the head of the caudate nucleus bilaterally in the Huntington's patients compared to healthy controls (projected on the mean grey matter segmentation from the 14 patients and 14 controls included in the study); left is on the left, $y = 10$ mm, $z = 10$ mm.

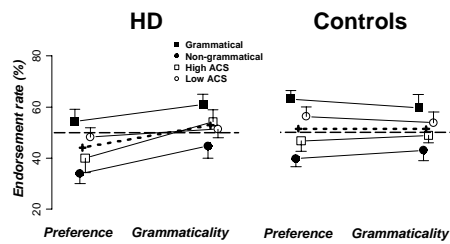


Figure 4.13 Preference and grammaticality classification performance (endorsement rate) for (a) main factors, and (b) factor levels; ACS = associative chunk strength.

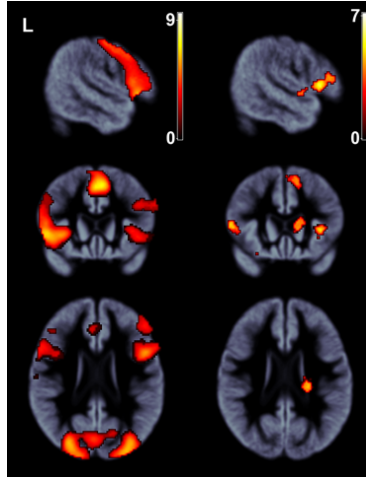


Figure 4.14 Regions significantly sensitive to grammatical classification (left) and regions significantly sensitive to performance, grammaticality, substring familiarity, or any interactions (omnibus ANOVA) in the Huntington group (right); corrected for multiple comparisons using the false discovery rate, projected on the study-specific mean grey matter segmentation, left is on the left, $x = -54$ mm, $y = 20$ mm, $z = 24$ mm.

CHAPTER 5 — CONCLUSIONS

We have in this thesis attempted to introduce the repeated artificial grammar learning model in the investigation of sequential implicit learning, in particular as a model for language acquisition and processing. We have however pointed out that the feature of implicit learning of sequential structure captures an essential cognitive processing capacity of interest from a larger cognitive neuroscience perspective. We set out in the objectives of this thesis (**section 1.1**) to investigate certain aspects of an extended version of the classical artificial grammar learning model, namely to

1. Explore the underlying neural processing architecture for implicit learning/acquisition to acquire and process non-motor sequences, an implicit non-motor procedural learning ability present in the human cognitive system (**section 1.3**),
2. Validate and explore the repeated artificial grammar learning paradigm as a laboratory model to investigate the acquisition and processing of structural aspects of language, e.g. (morpho-) syntax processing (**section 1.4**), and
3. Further our understanding of the specific neural processing architecture which subserves the language faculty and its syntax processing ability (**section 1.5**).

We present in this thesis a theoretical background on sequential procedural learning and formal grammars in cognitive processing (**section 1.3** and **1.4**) as well as give a general outline of the neuronal implementation of the cognitive functions involved (**section 1.5**). We suggest that as an alternative to describe implicit procedural acquisition as either rule acquisition or acquisition of statistical knowledge, it can be beneficial to take a lexical view on the processing and acquisition of artificial grammars, to understand the nature and representation of the acquired knowledge during artificial grammar learning. We pointed out that this pursuit re-traces a major trend in theoretical linguistics since the early 80's, so-called lexicalisation (cf., Culicover & Jackendoff, 2005; Jackendoff, 2002), in which the distinction between lexical items and grammatical rules is vanishing and in effect shifts more of the grammar into the mental lexicon. One lexical level we suggest to be of interest is the primitives which are the core of the formal language used in artificial grammars (**section 1.4**). From this perspective we suggest that formal grammar acquisition and processing of the (regular) grammar type commonly studied in artificial grammar learning can be used as a model to investigate the neuronal infrastructure supporting language acquisition and processing, including to characterize the neuronal infrastructure supporting syntax processing and unification (cf. e.g., Hagoort, 2003; Jackendoff, 1997; Jackendoff, 2007; Kaan & Swaab, 2002; Shieber, 1986; Vosse & Kempen, 2000).

In the studies included in this thesis we have managed to address some but not all of the objectives we set out to investigate. In **study 1** we began to describe the neuronal implementation using a setup based on the seminal study on implicit learning by Reber (1967). We found an overlap in the neural activation on artificial syntax violation and similar natural syntax violation. In **study 2** we replicated this finding using a more elaborated model with repeated acquisition sessions to simulate a prolonged acquisition period, and using a sequential presentation forcing the cognitive processing into a sequential processing mode. We found a clear advance for structural (grammaticality) knowledge over surface (ACS) knowledge where grammaticality increased over repeated acquisition sessions while ACS knowledge were stable over time. We also found a neuronal activation pattern which suggested that frontostriatal circuits were at play during artificial grammar classification, specifically the left inferior frontal region BA 44/45 and the head of the caudate nucleus. In this thesis we also report results from EEG measurements from the same study population, which suggests that the cognitive system can discriminate between grammatical and non-grammatical structures already on the first non-grammatical position in the sequence. In **study 3** we repeated the behaviour performance, introducing a preference classification instruction to further the cognitive system into an implicit learning mode to avoid explicit problem solving strategies during acquisition and classification, and found a clear and increasing preference for grammatical structure over repeated sessions, while the knowledge of ACS surface structure was stable or even diminished over sessions. In **study 4** we investigated the basal ganglia component in Huntington patients with specific caudate head lesions. We found that while the patients did not show any deficit in their behaviour performance, structures in the basal ganglia including the caudate head showed abnormal activation patterns compared to their matched normal controls. We also found a cooperative activation between basal ganglia and hippocampus typically involved in declarative memory. We interpreted this to reflect attempts of the cognitive system to compensate the damaged procedural processing with declarative knowledge processing. We conclude that we have in the studies of this thesis gained an initial characterization of the neural infrastructure subserving artificial grammar processing (**objective 1**) using FMRI, EEG, and alterations in behaviour performance. We have done so by characterising the end-state of the learning process (**study 1-4**), as well as characterizing the learning curves reflecting the outcome of acquisition at different time points (**study 2 and 3**). We further conclude that we have in this thesis reported findings supporting the view that the extended artificial grammar learning model is useful to capture structural aspects in language acquisition processing in the laboratory (**objective 2**). Finally, we conclude that we have in this thesis furthered our knowledge of the neuronal implementation and mechanisms at play in syntactic aspects of artificial language acquisition (**objective 3**), using implicit artificial grammar learning as a

method to engage brain regions involved in natural language processing. Further investigations will clarify the validity of this finding in natural language.

CHAPTER 6 — ACKNOWLEDGEMENTS

Inspiratörer:

Morfar Arne Fjelner

Trädklätraren Erik Engdahl

Kemi/Biologilärare, Heleneholmskolan

Theo van Veen

Margareta Ryberg

Jörgen Engel

Lars Svensson

Homan Alipour

Leif Nilsson

Martin Ingvar

Robert Fisk, John Pilger and other protagonists in the search of non-politically correct truths

Karl Magnus Petersson

Mina forskningssyskon Johan Sandblom, Jens Gisselgård, Katrina Carlsson och Predrag Petrovic.

Lorenzo Cangiano

The many young and the few senior researchers which have crossed my paths during my years at the Karolinska institute, the Donders Institute for Cognitive Neuroimaging and the Max Planck Institute for Psycholinguistics in the Netherlands, and at the Universidade do Algarve in Portugal.

Entusiaster:

Min Trygga Familj

The friends that crossed my path, took part of my life and sometimes kept in touch

The friends that still do

A minha Inês och min Camila, e a minha nova família Portuguesa

Mej själv.

CHAPTER 7 — REFERENCES

- Adler, Robert (1981). *The geometry of random fields*: Chichester : Wiley.
- Aguirre, Geoffrey, Zarahn, Eric, & D'Esposito, Mark (1997). Empirical analyses of BOLD fMRI statistics II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *NeuroImage*, 5, 199-212.
- Aguirre, Geoffrey, Zarahn, Eric, & D'Esposito, Mark (1998). The inferential impact of global signal covariates in functional neuroimaging analyses. *Neuroimage*, 8(3), 302-306.
- Anderson, John Robert (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cog. Sci.*, 26, 85-112.
- Arbib, Michael Anthony (Ed.). (2003). *The handbook of brain theory and neural networks* (2 ed.). Cambridge, MA: MIT Press.
- Arbib, Michael Anthony, & Rizzolatti, G (1997). Neural expectations: a possible evolutionary path from manual skills to language. *Communication and Cognition*, 29, 393-424.
- Ashburner, John, & Friston, Karl (1997). Multimodal image coregistration and partitioning - a unified framework. *Neuroimage*, 6(3), 209-217.
- Ashburner, John, & Friston, Karl (2000). Voxel-Based Morphometry - The Methods. *Neuroimage*, 11, 805-821.
- Ashburner, John, & Friston, Karl (2004). Computational Neuroanatomy. In Richard Frackowiak & Karl Friston & Christopher Frith & Raymond Dolan & Cathy Price & S Zeki & Ashburner J & Will Penny (Eds.), *Human Brain Function* (2 ed. ed., pp. 635-724). San Diego, CA: Academic Press.
- Ashburner, John, & Friston, Karl (2005). Unified segmentation. *Neuroimage*, 26(3), 839-851.
- Ashby, F G, Alfonso-Reese, L A, Turken, A U, & Waldron, E M (1998). A neuropsychological theory of multiple systems in category learning. *Psychol Rev*, 105(3), 442-481.
- Baddeley, Alan (1986). *Working memory*. Oxford: Clarendon press.
- Baddeley, Alan (1992). Working memory. *Science*, 255(5044), 556-559.
- Baddeley, Alan (2000). The episodic buffer: A new component of working memory? *Trends Cogn. Sci.*, 4, 417-423.
- Baddeley, Alan (2003). Working memory: looking back and looking forward. *Nat Rev Neurosci*, 4(10), 829-839.
- Baddeley, Alan, Gathercole, Susan, & Papagno, Constanza (1998). The phonological loop as a language learning device. *Psychological review*, 105(1), 158-173.
- Baddeley, Alan, & Hitch, Graham (1974). Working memory. In Gordon Bower (Ed.), *The Psychology of Learning and Motivation*. New York: Academic Press.

- Becker, J T, MacAndrew, D K, & Fiez, J A (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition*, 41, 27-38.
- Berwick, R C (1998). *Language evolution and the Minimalist Program: The origins of syntax*. In C Knight (Ed.), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 320-340). Cambridge, UK: Cambridge University Press.
- Bickel, P J, & Docksum, K A (1977). *Mathematical Statistics: Basic Ideas and Selected Topics* (1st ed.). Oakland, CA: Holden-Day.
- Bickerton, Derek (2000). How protolanguage became language. In Knight & Studdert-Kennedy & Hurford (Eds.), *The evolutionary emergence of language*.
- Binder, Marc D, Hirokawa, Nobutaka, & Windhorst, Uwe (Eds.) (2008). *Encyclopedia of Neuroscience*. New York: Springer.
- Bischoff-Grethe, A, Goedert, K M, Willingham, D T, & Grafton, S T (2004). Neural substrates of response-based sequence learning using fMRI. *J Cogn Neurosci*, 16, 127-138.
- Bitan, T, & Karni, A (2003). Alphabetical knowledge from whole words training: Effects of explicit instruction and implicit experience on learning script segmentation. *Cogn Brain Res*, 16, 325-339.
- Bitan, T, & Karni, A (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Brain Res Cogn Brain Res*, 19, 229-243.
- Bookheimer, S (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci*, 25, 151-188.
- Boyd, L A, & Winstein, C J (2004). Cerebellar stroke impairs temporal but not spatial accuracy during implicit motor learning. *Neurorehabil Neural Repair*, 18, 134-143.
- Breitenstein, C, & Knecht, S (2003). Language acquisition and statistical learning. *Nervenarzt*, 74(2), 133-143.
- Brett, M, Johnsrude, I S, & Owen, A M (2002). The problem of functional localization in the human brain. *Nat Rev Neurosci*, 3(3), 243-249.
- Broca, Paul (1861). *Remarques sur le siege de la faculte de langage articulation suivres d'une observation d'aphemie (perte de la parole)*. *Bulletine Societe d'Anatomie*, 36, 330-357.
- Buchner, Axel (1994). Indirect Effect of Synthetic Grammar Learning in an Identification task. *Journal of experimental psychology. Learning, memory, and cognition*, 20, 550-566.
- Buckner, R L, & Koutstaal, W (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proc Natl Acad Sci U S A*, 95(3), 891-898.
- Cabeza, R, Dolcos, F, Graham, R, & Nyberg, L (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage*, 16(2), 317-330.

- Cajochen, C, Knoblauch, V, Wirz-Justice, A, Krauchi, K, Graw, P, & Wallach, D (2004). Circadian modulation of sequence learning under high and low sleep pressure conditions. *Behav Brain Res*, 151, 167-176.
- Caplan, D, Alpert, N, & Waters, G (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *J Cogn Neurosci*, 10(4), 541-552.
- Caplan, D, Alpert, N, & Waters, G (1999). PET studies of syntactic processing with auditory sentence presentation. *Neuroimage*, 9(3), 343-351.
- Caplan, D, Alpert, N, Waters, G, & Olivieri, A (2000). Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Hum Brain Mapp*, 9(2), 65-71.
- Caplan, David (1992). *Language : structure, processing, and disorders*: Cambridge, Mass. : MIT Press.
- Caramazza, A, & Zurif, E B (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain Lang*, 3(4), 572-582.
- Chang, G Y, & Knowlton, B J (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 30, 714-722.
- Channon, S, Shanks, D, Johnstone, T, Vakili, K, Chin, J, & Sinclair, E (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, 40(12), 2185-2197.
- Charniak, E (1993). *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Charniak, E, & McDermott, D (1985). *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley.
- Cherkassky, Vladimir, & Mulier, Filip M (1998). *Learning from data: concepts, theory and methods*: New York ; Chichester : Wiley.
- Chomsky, Noam (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.
- Chomsky, Noam (1959). Review of Skinner's *Verbal Behavior*. *Language*, 35, 26-58.
- Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam (1981). *Lectures on government and binding: the Pisa lectures*: Dordrecht, The netherlands: Foris publications.
- Chomsky, Noam (1986). *Knowledge of Language*. New York: Praeger.
- Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam (2000). *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.
- Chomsky, Noam, & Lasnik, H (1995). *The Theory of Principles and Parameters*. In N Chomsky (Ed.), *The Minimalist Program* (pp. 13-128). Cambridge, MA: MIT Press.
- Christiansen, M H, & Chater, N (Eds.) (2001). *Connectionist Psycholinguistics*. Norwood, NJ: Ablex Publishing.

- Christiansen, Morten H, & Chater, Nick (1999). *Toward a connectionist model of recursion in human linguistic performance. Cognitive Science: A Multidisciplinary Journal*, 23(2), 157-205.
- Christie, M A, & Dalrymple-Alford, J C (2004). *A new rat model of the human serial reaction time task: Contrasting effects of caudate and hippocampal lesions. J Neurosci*, 24, 1034-1039.
- Churchland, Patricia S, & Sejnowski, Terrence J (1993). *The computational brain. Cambridge: MIT Press.*
- Cleeremans, A, & McClelland, J L (1991). *Learning the structure of event sequences. J Exp Psychol Gen*, 120(3), 235-253.
- Cohen, D I A (1997). *Introduction to Computer Theory (2nd ed.). New York: Wiley and Sons.*
- Cohen, J D, Barch, D M, Carter, Cameron, & Servan-Schreiber, D (1999). *Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. J Abnorm Psychol*, 108(1), 120-133.
- Cohen, J D, & Servan-Schreiber, D (1992). *Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol Rev*, 99(1), 45-77.
- Cohen, J D, Dunbar, K, & McClelland, J L (1990). *On the control of automatic processes: A parallel distributed processing model of the Stroop effect. Psychol. Rev.*, 99, 45-77.
- Cohen, N J, & Squire, L R (1980). *Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science*, 210(4466), 207-210.
- Conway, C M, & Christiansen, M H (2001). *Sequential learning in non-human primates. Trends Cogn Sci*, 5(12), 539-546.
- Cooke, A, Zurif, E B, DeVita, C, Alsop, D, Koenig, P, Detre, J, Gee, J, Pinango, M, Balogh, J, & Grossman, M (2001). *Neural basis for sentence comprehension: Grammatical and short-term memory components. Hum. Brain Map.*, 15, 80-94.
- Cooper, Jack R, Bloom, Floyd E, & Roth, Robert H (1996). *The Biochemical Basis of Neuropharmacology (Seventh edition ed.). New York: Oxford University Press.*
- Corballis, M C (2003). *From hand to mouth: The origins of language. Princeton, NJ: Princeton University Press.*
- Culicover, Peter W, & Jackendoff, Ray (2005). *Simpler Syntax. Oxford, UK: Oxford University Press.*
- Curran, T (1998). *Implicit sequence learning from a cognitive neuroscience perspective. In M A Stadler & P A Frensch (Eds.), Handbook of Implicit Learning. Thousand Oaks, CA: Sage.*
- Cutland, N J (1980). *Computability: An Introduction to Recursive Function Theory. Cambridge, UK: Cambridge University Press.*

- Dapretto, M, & Bookheimer, SY (1999). *Form and content: dissociating syntax and semantics in sentence comprehension.* *Neuron*, 24(427-432).
- Daselaar, S M, Rombouts, S A, Veltman, D J, Raaijmakers, J G, & Jonker, C (2003). *Similar network activated by young and old adults during the acquisition of a motor sequence.* *Neurobiol Aging*, 24, 1013-1019.
- Davis, M D, Sigal, R, & Weyuker, E J (1994). *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science (2 ed.).* San Diego, CA: Academic Press.
- Deacon, Terrence W (1997). *The symbolic species: The pinguine press.*
- DeCoteau, W E, & Kesner, R P (2000). *A double dissociation between the rat hippocampus and medial caudoputamen in processing two forms of knowledge.* *Behav Neurosci*, 114, 1096–1108.
- Dehaene, S, Kerszberg, M, & Changeux, J P (1998). *A neuronal model of a global workspace in effortful cognitive tasks.* *Proc Natl Acad Sci U S A*, 95(24), 14529-14534.
- Destrebecqz, Arnaud, & Cleeremans, Axel (2001). *Can sequence learning be implicit? New evidence with the process dissociation procedure.* *Psychonomic Bulletin & Review*, 8(2) Jun 2001.
- Dienes, Z, & Perner, J (1999). *A theory of implicit and explicit knowledge.* *Behav Brain Sci*, 22(5), 735-755; discussion 755-808.
- Domangue, T J, Mathews, R C, Sun, R, Roussel, L G, & Guidry, C E (2004). *Effects of model-based and memory-based processing on speed and accuracy of grammar string generation.* *J Exp Psychol Learn Mem Cogn*, 30(5), 1002-1011.
- Doyon, J, Owen, A M, Petrides, M, Sziklas, V, & Evans, A C (1996). *Functional anatomy of visuomotor skill learning in human subjects examined with positron emission tomography.* *The European Journal of Neuroscience*, 8(4), 637-648.
- Doyon, J, Penhune, V, & Ungerleider, L G (2003). *Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning.* *Neuropsychologia*, 41, 252-262.
- Dronkers, N F (2000). *The pursuit of brain–language relationships.* *Brain and Language*, 71, 59-61.
- Duda, R O, Hart, P E, & Stork, D G (2001). *Pattern Classification (2nd ed.).* New York: Wiley and Sons.
- Duncan, J (2001). *An adaptive coding model of neural function in prefrontal cortex.* *Nat Rev Neurosci*, 2(11), 820-829.
- Duncan, J, & Miller, E K (2002). *Cognitive focus through adaptive neural coding in the primate prefrontal cortex.* In R T Knight (Ed.), *Principles of Frontal Lobe Function* (pp. 278-292). Oxford, UK: Oxford University Press.

- Edelman, G M (1990). *Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- Edgington, E S (1995). *Randomization Tests (Third, revised and expanded ed.)*. New York: Marcel Dekker.
- Eichenbaum, H (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.*, 1, 41-50.
- Eichenbaum, H, & Cohen, N J (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. Oxford, UK: Oxford University Press.
- Ellefson, Michelle R, & Christiansen, Morten H (2000). Subjacency constraints without universal grammar: evidence from artificial language learning and connectionist modeling. *Proc Cogn Sci Soc*, 645-650.
- Elman, J L (1990). Finding structure in time. *Cogn. Sci.*, 14, 179-211.
- Elman, J L, Bates, E A, Johnson, M H, Karmiloff-Smith, A, Parisi, D, & Plunkett, K (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Embick, D, Marantz, A, Miyashita, Y, O'Neil, W, & Sakai, K L (2000). A syntactic specialization for Broca's area. *Proc. Natl. Acad. Sci. USA*, 97, 6150-6154.
- Engel, A, & Van den Broeck, C (2001). *Statistical Mechanics of Learning*. Cambridge, UK: Cambridge University Press.
- Epstein, S (1999). Un-principled syntax and the derivation of syntactic relations. In N Hornstein (Ed.), *Working Minimalism*. Cambridge, MA: MIT Press.
- Felleman, D J, & Van Essen, D C (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1), 1-47.
- Fletcher, P, Buchel, C, Josephs, O, Friston, Karl, & Dolan, Raymond (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cereb Cortex*, 9(2), 168-178.
- Fletcher, P C, & Henson, R N A (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain*, 124, 849-881.
- Fletcher, P C, Zafiris, O, Frith, Christiapher, Honey, R A, Corlett, P R, Zilles, Karl, & Fink, G R (2004). On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cereb Cortex*.
- Fodor, J (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.
- Fodor, J A, & Pylyshyn, Z W (1990). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Fodor, Jerry A (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge, Mass.: MIT Press.

- Folia, Vasiliki, Uddén, Julia, Forkstam, Christian, Ingvar, Martin, Hagoort, Peter, & Petersson, Karl Magnus (2008). *Implicit learning and dyslexia*. *Ann N Y Acad Sci*, 1145, 132-150.
- Forkstam, Christian, Elwér, Åsa, Ingvar, Martin, & Petersson, Karl Magnus (2008). *Instruction effects in implicit artificial grammar learning: A preference for grammaticality*. *Brain Res*, 1221, 80-92.
- Forkstam, Christian, Hagoort, Peter, Fernandez, Guillen, Ingvar, Martin, & Petersson, Karl Magnus (2006). *Neural correlates of artificial syntactic structure classification*. *NeuroImage*, 32(2), 956-967.
- Forkstam, Christian, & Petersson, Karl Magnus (2005). *Towards an explicit account of implicit learning*. *Curr Opin Neurol*, 18(4), 435-441.
- Forkstam, Christian, Voermans, Nicol, Dekkers, Marieke, Kremer, Berry, Fernández, Guillen, & Petersson, Karl Magnus (In submission). *Frontostriatal circuitry in artificial grammar learning: An FMRI study in Huntington's disease*.
- Fox, P T, & Pardo, J V (1991). *Does inter-subject variability in cortical functional organization increase with neural 'distance' from the periphery?* *Ciba Found Symp*, 163, 125-140; discussion 140-124.
- Fox, P T, Raichle, M E, Mintun, M A, & Dence, C (1988). *Nonoxidative glucose consumption during focal physiologic neural activity*. *Science*, 241(4864), 462-464.
- Frackowiak, Richard, Friston, Karl, Frith, Christopher, Dolan, Raymond, Price, Cathy, Zeki, S, Ashburner, John, & Penny, Will (Eds.) (2004). *Human brain function (2 ed.)*. San Diego, CA: Academic Press.
- French, Robert M, & Cleeremans, Axel (Eds.) (2002). *Implicit Learning and Consciousness*. Hove, UK: Psychology Press.
- Friederici, A D, Steinhauer, K, & Pfeifer, E (2002). *Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis*. *Proc Natl Acad Sci U S A*, 99, 529-534.
- Friederici, Angela D (2002). *Towards a neural basis of auditory sentence processing*. *Trends in Cognitive Sciences*, 6(2), 78-84.
- Friston, Karl, Harrison, L, & Penny, Will (2003). *Dynamic causal modelling*. *Neuroimage*, 19(4), 1273-1302.
- 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). *The Prefrontal Cortex: Anatomy, physiology, and neuropsychology of the frontal lobe (3 ed ed.)*. New York: Lippincott-Raven.
- Gabrieli, J D (1998). *Cognitive neuroscience of human memory*. *Annu Rev Psychol*, 49, 87-115.

- Gagnon, S, Foster, J K, Turcotte, J, & Jongenelis, S (2004). *Involvement of the hippocampus in implicit learning of supra-span sequences: The case of SJ*. *Cognitive Neuropsychology*, 21, 867-882.
- Gazzaniga, M S (Ed.). (1995). *The cognitive neurosciences*. Cambridge, MA: MIT press.
- Gazzaniga, Michael S (Ed.). (2000). *The new cognitive neurosciences*. Cambridge, Mass.: MIT Press.
- Gazzaniga, Michael S (Ed.). (2004). *The Cognitive Neurosciences III (3ed ed.)*. Cambridge, Mass.: MIT Press.
- Gelfand, J R, & Bookheimer, S Y (2003). *Dissociating neural mechanisms of temporal sequencing and processing phonemes*. *Neuron*, 38, 831-842.
- Gerstner, W, & Kistler, W (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: Cambridge University Press.
- Gold, EM (1967). *Language identification in the limit*. *Information and Control*, 10, 447-474.
- Goldman-Rakic, P S (1988). *Topography of cognition: parallel distributed networks in primate association cortex*. *Annu Rev Neurosci*, 11, 137-156.
- Goldman-Rakic, PS (1998). *The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive*. In A C Roberts & T W Robbins & L Weiskrantz (Eds.), *The Prefrontal Cortex: Executive and Cognitive Functions* (pp. 87-102). Oxford, UK: Oxford University Press.
- Gomez, Rebecca Leanne, & Gerken, L (1999). *Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge*. *Cognition*, 70, 109-135.
- Gomez, Rebecca Leanne, & Gerken, L (2000). *Infant artificial language learning and language acquisition*. *Trends in Cognitive Sciences*, 4, 178-186.
- Gomez, Rebecca Leanne, & Schvaneveldt, Roger W (1994). *What is learned from artificial grammars? Transfer tests of simple association*. *J Exp Psychol Learn Mem Cogn*, 20(2), 396-410.
- Good, C D, Scahill, R I, Fox, N C, Ashburner, John, Friston, Karl, Chan, D, Crum, W R, Rossor, M N, & Frackowiak, Richard (2002). *Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias*. *Neuroimage*, 17(1), 29-46.
- Grabowski, T J, Frank, R J, Brown, C K, Damasio, H, Boles Ponto, L L, Watkins, G L, & Hichwa, R D (1996). *Reliability of PET activation across statistical methods, subject groups, and sample sizes*. *Human Brain Mapping*, 4(1), 23-46.
- Grafton, Scott T, Hazeltine, Eliot, & Ivry, Richard (1995). *Functional mapping of sequence learning in normal humans*. *Journal of cognitive neuroscience*, 7(4), 497-510.
- Gregory, R L (1961). *The brain as an engineering problem*. In W H Thorpe & O L Zangwill (Eds.), *Current Problems in Animal Behaviour: Methuen*.

- Grossberg, S (1999). *The link between brain learning, attention, and consciousness*. *Conscious Cogn*, 8(1), 1-44.
- Gusnard, D A, Raichle, M E, & Raichle, M E (2001). *Searching for a baseline: functional imaging and the resting human brain*. *Nat Rev Neurosci*, 2(10), 685-694.
- Haber, S N, & Gdowski, M J (2004). *The basal ganglia*. In G Paxinos & J K Mai (Eds.), *The human nervous system* (2nd ed., pp. 676–738). New York, NY: Elsevier, Academic Press.
- 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, Peter (2003). *How the brain solves the binding problem for language: a neurocomputational model of syntactic processing*. *Neuroimage*, 20 Suppl 1(S18-29), S18-29.
- Hagoort, Peter (2004). *Broca's complex as the unification space for language*. In Anne Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (Vol. 20, pp. s18-s29): Erlbaum.
- Hagoort, Peter (2005). *On Broca, brain and binding: A new framework*. *Trends Cog. Sci.*, 9, 416-423.
- Harlow, H F (1949). *The formation of learning sets*. *Psychol Rev*, 56(1), 51-65.
- Hasnain, M K, Fox, P T, & Woldorff, M G (1998). *Intersubject variability of functional areas in the human visual cortex*. *Hum Brain Mapp*, 6(4), 301-315.
- Hauser, M D, & McDermott, J (2003). *The evolution of the music faculty: a comparative perspective*. *Nat Neurosci*, 6(7), 663-668.
- Hauser, Marc D, Chomsky, Noam, & Fitch, W Tecumseh (2002). *The faculty of language: what is it, who has it, and how did it evolve?* *Science*, 298, 1569-1579.
- Haykin, S (1994). *Neural networks: a comprehensive foundation* (1st ed.). NY: Macmillan.
- Haykin, S (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Hebb, Donald O (1949). *Organization of Behavior*. New York: Wiley.
- Hendelman, Walter (2006). *Atlas of Functional Neuroanatomy* (2ed ed.): CRC Press.
- Hochberg, Y, & Tamhane, A C (1987). *Multiple Comparisons Procedures*. New York: Wiley and Sons.
- Hopcroft, J E, Motwani, R, & Ullman, J D (2000). *Introduction to Automata Theory, Languages, and Computation* (2nd ed.). Reading, MA: Addison Wesley.
- Hopkins, R O, Waldram, K, & Kesner, R P (2004). *Sequences assessed by declarative and procedural tests of memory in amnesic patients with hippocampal damage*. *Neuropsychologia*, 42, 1877-1886.

- Horgan, Terence, & Tienson, John (1996). *Connectionism and the philosophy of psychology*. Cambridge, Mass.: MIT Press.
- Hunton, D L, Miezin, F M, Buckner, R L, van Mier, H I, Raichle, M E, & Petersen, S E (1996). An assessment of functional-anatomical variability in neuroimaging studies. *Human Brain Mapping*, 4, 122-139.
- Indefrey, P (2004). Hirnaktivierungen bei syntaktischer Sprachverarbeitung: eine Meta-Analyse. In H M Müller & G Rickheit (Eds.), *Neurokognition der Sprache* (pp. 31-50). Tuebingen: Stauffenburg Verlag.
- 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(3), 546-555.
- Inui, T, Otsu, Y, Tanaka, S, Okada, T, Nishizawa, S, & Konishi, J (1998). A functional MRI analysis of comprehension processes of Japanese sentences. *NeuroReport*, 9, 3325-3328.
- Jackendoff, R (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Jackendoff, Ray (2002). *Foundations of language: Brain, Meaning, Grammar, Evolution*. Oxford, UK: Oxford University Press.
- Jackendoff, Ray (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2-22.
- Jacoby, Larry L (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of memory and language*, 30, 513-541.
- Jain, S, Osherson, D, Royer, J S, & Sharma, A (1999). *Systems That Learn*. Cambridge, MA: MIT Press.
- Janata, Petr, & Grafton, Scott T (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature neuroscience*, 6(7), 682-687.
- Jimenez, L, & Mendez, C (1999). Which attention is needed for implicit sequence learning? *J Exp Psychol Learn Mem Cogn*, 25, 236-259.
- Johnstone, T, & Shanks, D R (2001). Abstractionist and processing accounts of implicit learning. *Cognit Psychol*, 42(1), 61-112.
- Joshi, A K, & Schabes, Y (1997). *Tree-adjoining grammars* (Vol. 3). Berlin: Springer Verlag.
- Just, MA, Carpenter, PA, Keller, TA, Eddy, WF, & Thulborn, KR (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114-116.
- Kaan, E, & Swaab, T Y (2002). The brain circuitry of syntactic comprehension. *Trends Cogn Sci*, 6(8), 350-356.
- Kager, R (1999). *Optimality Theory*. Cambridge, UK: Cambridge University Press.
- Kang, A M, Constable, R T, Gore, J C, & Avrutin, S (1999). An event-related fMRI study of implicit phrase-level syntactic and semantic processing. *Neuroimage*, 10(5), 555-561.

- Kelly, S D, Ward, S, Creigh, P, & Bartolotti, J (2007). *An intentional stance modulates the integration of gesture and speech during comprehension. Brain Lang, 101(3), 222-233.*
- Kelly, S W, Jahanshahi, M, & Dirnberger, G (2004). *Learning of ambiguous versus hybrid sequences by patients with Parkinson's disease. Neuropsychologia, 42, 1350-1357.*
- Kim, J S, Reading, S A, Brashers-Krug, T, Calhoun, V D, Ross, C A, & Pearlson, G D (2004). *Functional MRI study of a serial reaction time task in Huntington's disease. Psychiatry Res, 131(1), 23-30.*
- Kluender, R, & Kutas, M (1993). *Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. Journal of Cognitive Neuroscience(5), 196-214.*
- Knowlton, B J, Mangels, J A, & Squire, L R (1996a). *A neostriatal habit learning system in humans. Science, 273, 1399-1401.*
- Knowlton, B J, & Squire, L R (1996). *Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. J Exp Psychol Learn Mem Cogn, 22, 169-181.*
- Knowlton, Barbara J, & Squire, Larry R (1997). "Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information": Correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(1), 220.*
- Knowlton, Barbara J, Squire, Larry R, Paulsen, Jane S, Swerdlow, Neal R, & Swenson, Michael (1996b). *Dissociations within nondeclarative memory in Huntington's disease. Neuropsychology, 10, 538-548.*
- Koch, C (1999). *Biophysics of Computation: Information Processing in Single Neurons. New York: Oxford University Press.*
- Koch, C, & Laurent, G (1999). *Complexity and the nervous system. Science, 284(5411), 96-98.*
- Koch, C, & Segev, I (1998). *Methods in Neuronal Modeling: From Ions to Networks (2nd ed.). New York: Oxford University Press.*
- Kutas, M, & Hillyard, S A (1980). *Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207(207), 203-205.*
- Lasnik, H (2000). *Syntactic Structures Revisited: Contemporary Lectures on Classic Transformational Theory. Cambridge, MA: MIT Press.*
- Lelekov, T, Dominey, P F, & Garcia-Larrea, L (2000). *Dissociable ERP profiles for processing rules vs instances in a cognitive sequencing task. Neuroreport, 11(5), 1129-1132.*
- Lerdahl, Fred, & Jackendoff, Ray (1983). *A generative theory of tonal music. Cambridge, Mass: MIT Press.*
- Lewicki, P (1986). *Nonconscious social information processing. San Diego: Academic Press.*
- Lewis, H R, & Papadimitriou, C H (1981). *Elements of the Theory of Computation. Englewood Cliffs, NJ: Prentice-Hall.*

- Lieberman, M D, Chang, G Y, Chiao, J, Bookheimer, S Y, & Knowlton, B J (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J Cogn Neurosci*, 16(3), 427-438.
- Logothetis, N K, Guggenberger, H, Peled, S, & Pauls, J (1999). Functional imaging of the monkey brain. *Nat Neurosci*, 2(6), 555-562.
- Logothetis, N K, Pauls, J, Augath, M, Trinath, T, & Oeltermann, A (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150-157.
- Lungu, O V, Wachter, T, Liu, T, Willingham, D T, & Ashe, J (2004). Probability detection mechanisms and motor learning. *Exp Brain Res*, 159, 135-150.
- Macmillan, Neil A, & Creelman, C Douglas (1991). *Detection theory: a user's guide*. Cambridge: Cambridge University Press.
- Manning, C D, & Schütze, H (2000). *Foundations of Statistical Natural Language Processing, 2nd printing with corrections*. Cambridge, MA: MIT Press.
- Manza, L, & Bornstein, R F (1995). Affective discrimination and the implicit learning process. *Consciousness and Cognition*, 4, 399-409.
- Maquet, P (2001). The role of sleep in learning and memory. *Science*, 294, 1048-1052.
- Marcus, G F, Vijayan, S, Bandi Rao, S, & Vishton, P M (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-80.
- Marcus, G F, Vouloumanos, A, & Sag, I A (2003). Does Broca's play by the rules? *Nat Neurosci.*, 6, 651-652.
- Maris, Eric (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41(1), 142-151.
- Marr, D (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.
- McClelland, J L (1994). The organization of memory. A parallel distributed processing perspective. *Rev Neurol (Paris)*, 150(8-9), 570-579.
- McCulloch, W S, & Pitts, W (1943). A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*, 5, 115-133.
- McIntosh, A R, & Gonzalez-Lima, F (1994). Structural equation modeling and its application to network analysis in functional brain imaging (Vol. 2, pp. 2-22).
- Menon, V, Ford, J M, Lim, K O, Glover, G H, & Pfefferbaum, A (1997). Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport*, 8(14), 3029-3037.
- Mesulam, M M (Ed.). (2002). *The human frontal lobes: Transcending the default mode through contingent encoding*. Oxford, UK: Oxford University Press.
- Meulemans, Thierry, & Van der Linden, Martial (1997). Associative chunk strength in artificial grammar learning. *J Exp Psychol Learn Mem Cogn*, 23, 1007-1028.

- Miller, E K, & Cohen, J D (2001). *An integrative theory of prefrontal cortex function. Annu Rev Neurosci*, 24, 167-202.
- Minsky, M L (1967). *Computation: Finite and Infinite Machines. Englewood Cliffs, NJ: Prentice-Hall.*
- Musso, Mariacristina, Moro, Andrea, Glauche, Volkmar, Rijntjes, Michel, Reichenbach, Jürgen, Büchel, Christian, & Weiller, Cornelius (2003). *Broca's area and the language instinct. Nature neuroscience*, 6(7), 774-781.
- Newell, A (1990). *Unified Theories of Cognition: Harvard University Press.*
- Newell, Allen (1992). *Precis of Unified Theories of Cognition. Behavioral and Brain Sciences*, 15(3), 425-492.
- Newell, Ben R, & Bright, James E (2001). *The relationship between the structural mere exposure effect and the implicit learning process. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 54A(4), 1087-1104.
- Newmeyer, FJ (1995). *Generative linguistics. New York: Routledge.*
- Ni, W, Constable, R T, Mencl, W E, Pugh, K R, Fulbright, R K, Shaywitz, S E, Shaywitz, B A, Gore, J C, & Shankweiler, D (2000). *An event-related neuroimaging study distinguishing form and content in sentence processing. J Cogn Neurosci*, 12(1), 120-133.
- Niedeggen, M., Rosler, F., & Jost, K. (1999). *Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. Psychophysiology*, 36(3), 307-324.
- Nieuwenhuis, Ingrid, Folia, Vasiliki, Forkstam, Christian, Jensen, Ole, & Petersson, Karl Magnus (In submission). *Grammar learning requires sleep.*
- Nieuwenhuys, Rudolf, Voogd, Jan, & van Huijzen, Christiaan (1978). *The human central system: a synopsis and atlas.*
- Nissen, Mary Jo, & Bullemer, Peter (1987). *Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology*, 19(1), 1-32.
- Nowak, Martin A, Komarova, Natalla L, & Niyogi, Partha (2002). *Computational and evolutionary aspects of language. Nature*, 417, 611-617.
- Nyberg, Lars, Forkstam, Christian, Petersson, Karl Magnus, Cabeza, Roberto, & Ingvar, Martin (2002). *Brain imaging of human memory systems: between-systems similarities and within-system differences. Brain Res Cogn Brain Res*, 13(2), 281-292.
- Nyberg, Lars, Marklund, Peter, Persson, J, Cabeza, R, Forkstam, Christian, Petersson, Karl Magnus, & Ingvar, Martin (2003). *Common prefrontal activations during working memory, episodic memory, and semantic memory. Neuropsychologia*, 41(3), 371-377.
- Ojemann, G A (1983). *Brain organization for language from the perspective of electrical stimulation mapping. Behav. Brain Sci.*, 6, 189-230.

- Ojemann, G A, & Mateer, C (1979). *Human language cortex: Localization of memory, syntax, and sequential motor-phoneme identification systems. Science, 205, 1401-1403.*
- Opitz, B, & Friederici, A D (2003). *Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. Neuroimage, 19(4), 1730-1737.*
- Opitz, B, & Friederici, A D (2004). *Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. J Neurosci, 24(39), 8436-8440.*
- Opitz, B, Mecklinger, A, Friederici, A D, & von Cramon, D Y (1999). *The functional neuroanatomy of novelty processing: integrating ERP and fMRI results. Cereb Cortex, 9(4), 379-391.*
- Osterhout, L, & Holcomb, P J (1992). *Event-Related Brain Potentials Elicited by Syntactic Anomaly. J Mem Lang, 31, 785-806.*
- Partee, BH, ter Meulen, A, & Wall, RE (1990). *Mathematical methods in linguistics. Dordrecht, the Netherlands: Kluwer academic publishers.*
- Pascual-Leone, A, Bartres-Faz, D, & Keenan, J P (1999). *Transcranial magnetic stimulation: studying the brain-behaviour relationship by induction of 'virtual lesions' [In Process Citation]. Philos Trans R Soc Lond B Biol Sci, 354(1387), 1229-1238.*
- Passingham, R E, Stephan, K E, & Kotter, R (2002). *The anatomical basis of functional localization in the cortex. Nat Rev Neurosci, 3(8), 606-616.*
- Patel, A D (2003). *Language, music, syntax and the brain. Nat Neurosci, 6(7), 674-681.*
- Patel, A D, Gibson, E, Ratner, J, Besson, M, & Holcomb, P J (1998). *Processing syntactic relations in language and music: An event-related potential study. J. Cogn. Neurosci., 10, 717-733.*
- Paterson, S J, Brown, J H, Gsodl, M K, Johnson, M H, & Karmiloff-Smith, A (1999). *Cognitive modularity and genetic disorders. Science, 286(5448), 2355-2358.*
- Paulesu, E, Frith, Christopher, & Frackowiak, Richard (1993). *The neural correlates of the verbal component of working memory. Nature, 362, 342-345.*
- Peigneux, P, Meulemans, T, Van der Linden, M, Salmon, E, & Petit, H (1999). *Exploration of implicit artificial grammar learning in Parkinson's disease. Acta Neurol Belg, 99(2), 107-117.*
- Peretz, I, & Coltheart, M (2003). *Modularity of music processing. Nat Neurosci, 6(7), 688-691.*
- Perruchet, Pierre, & Pacteau, Chantal (1991). *Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. Journal of experimental psychology. General, 120, 112-116.*
- Petersson, Karl Magnus (2005a). *Learning and Memory in the Human Brain. Karolinska University Press, Stockholm, Sweden.*

- Petersson, Karl Magnus (2005b). *On the relevance of the neurobiological analogue of the finite state machine*. *Neurocomputing*, 65-66, 825-832.
- Petersson, Karl Magnus, Forkstam, Christian, & Ingvar, Martin (2004). *Artificial syntactic violations activate Broca's region*. *Cognitive Science*, 28, 383-407.
- Petersson, Karl Magnus, Grenholm, Peter, & Forkstam, Christian (2005). *Artificial grammar learning and neural networks*. *Proceedings of the Cognitive Science Society*, 1726-1731.
- Petersson, Karl Magnus, Ingvar, Martin, & Reis, Alexandra (2009). *Language and literacy from a cognitive neuroscience perspective*. In D R Olson & N Torrence (Eds.), *Cambridge Handbook of Literacy*. New York: Cambridge University Press.
- Petersson, Karl Magnus, Nichols, T E, Poline, J -B, & Holmes, A P (1999a). *Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models*. *Philos Trans R Soc Lond B Biol Sci*, 354(1387), 1239-1260.
- Petersson, Karl Magnus, Nichols, T E, Poline, J -B, & Holmes, A P (1999b). *Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference*. *Philos Trans R Soc Lond B Biol Sci*, 354(1387), 1261-1281.
- Petrides, M (1995). *Functional organization of the human frontal cortex for mnemonic processing. Evidence from neuroimaging studies*. *Ann. N. Y. Acad. Sci.*, 769, 85-96.
- Pinker, S, & Jackendoff, R (2005). *The faculty of language: what's special about it?* *Cognition*, 95(2), 201-236.
- Pinker, Steven (1994). *The language instinct*. New York, NY: William Morrow & Co.
- Plaut, D C (1995). *Double dissociation without modularity: Evidence from connectionist neuropsychology*. *J. Clin. Exp. Neuropsychol.*, 17, 291-331.
- Poizner, H, Bellugi, U, & Klima, E S (1991). *Brain function for language: Perspectives from another modality*. In I Mattingly & M Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 145-170). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Poletiek, F H (2002). *Implicit learning of a recursive rule in an artificial grammar*. *Acta psychologica*, 111, 323-335.
- Poline, J B, Vandenberghe, R, Holmes, A P, Friston, Karl, & Frackowiak, Richard (1996). *Reproducibility of PET activation studies: lessons from a multi-center European experiment*. *EU concerted action on functional imaging*. *Neuroimage*, 4(1), 34-54.
- Posner, M I (Ed.). (1989). *The Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- Posner, M I, & Raichle, M E (1994). *Images of mind*. NY: New York, N.Y. : Scientific American Library.
- Pullum, G K, & Scholz, B C (2002). *Empirical assessment of stimulus poverty arguments*. *Ling. Rev.*, 19, 9-50.

- Quine, W V (1992). *Pursuit of Truth*: Harvard University Press.
- Radford, A (1997). *Syntax: A Minimalist Introduction*. Cambridge, UK: Cambridge University Press.
- Radford, A (2000). *Children in search of perfection: Towards a minimalist model of language acquisition*. *Essex Research Reports in Linguistics*, 34, 57-74.
- Raichle, M E (1994). *Images of the mind: studies with modern imaging techniques*. *Annu Rev Psychol*, 45, 333-356.
- Raichle, M E (1997). *Food for thought. The metabolic and circulatory requirements of cognition*. *Ann N Y Acad Sci*, 835, 373-385.
- Raichle, M E (2001). *Cognitive neuroscience. Bold insights*. *Nature*, 412(6843), 128-130.
- Ramsey, N F, Kirkby, B S, Van Gelderen, P, Berman, K F, Duyn, J H, Frank, J A, Mattay, V S, Van Horn, J D, Esposito, G, Moonen, C T, & Weinberger, D R (1996). *Functional mapping of human sensorimotor cortex with 3D BOLD fMRI correlates highly with H2(15)O PET rCBF*. *J Cereb Blood Flow Metab*, 16(5), 755-764.
- Rauch, S L, Whalen, P J, Savage, C R, Curran, T, Kendrick, A, Brown, H D, Bush, G, Breiter, H C, & Rosen, B R (1997). *Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging*. *Hum Brain Mapp*, 5(2), 124-132.
- Reber, Arthur (1967). *Implicit learning of artificial grammars*. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Reber, Arthur (1989). *Implicit learning and tacit knowledge*. *Journal of Experimental Psychology: General*, 118(3), 219-235.
- Reber, Arthur (1993). *Implicit learning and tacit knowledge: an essay on the cognitive unconscious*. New York, NY: Oxford University Press.
- Reber, Paul, & Squire, L R (1994). *Parallel brain systems for learning with and without awareness*. *Learn Mem*, 1, 217-229.
- Reber, Paul, & Squire, Larry (1998). *Encapsulation of Implicit and Explicit Memory in Sequence Learning*. *Journal of Cognitive Neuroscience*, 10(2), 248-263.
- Redington, M, & Chater, N (2002). *Knowledge representation and transfer in artificial grammar learning*. In R French & A Cleeremans (Eds.), *Implicit learning and consciousness* (pp. 121-143). Hove, UK: Psychology Press.
- Redington, Martin, & Chater, Nick (1996). *Transfer in artificial grammar learning: A reevaluation*. *Journal of Experimental Psychology: General*, 125(2), 123-138.
- Reed, Jonathan, & Johnson, Peder (1994). *Assessing implicit learning with indirect tests: Determining what is learned about sequence structure*. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20(3), 585-594.

- Rees, G, Friston, Karl, & Koch, C (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nat Neurosci*, 3(7), 716-723.
- Reingold, E M, & Merikle, P M (1988). Using direct and indirect measures to study perception without awareness. *Percept Psychophys*, 44(6), 563-575.
- Rieke, F, Warland, D, van Steveninck, R R D, & Bialek, W (1996). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.
- Rizzolatti, G, & Arbib, Michael Anthony (1998). Language within our grasp. *Trends Neurosci*, 21(5), 188-194.
- Robertson, E M (2004). Skill learning: putting procedural consolidation in context. *Curr Biol*, 14, R1061-R1063.
- Robertson, E M, Pascual-Leone, A, & Press, D Z (2004). Awareness modifies the skill-learning benefits of sleep. *Curr Biol*, 14, 208-212.
- Rogers, H (2002). *Theory of Recursive Functions and Effective Computability (5th ed.)*. Cambridge, MA: MIT Press.
- Rohde, Douglas, & Plaut, David (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition*, 72, 67-109.
- Rosenbleuth, A, Wiener, N, & Bigelow, J (1943). Behavior, purpose, and teleology. *Philos. Sci.*, 10, 18-24.
- Rosenfeld, A, & Kak, A C (1982). *Digital Picture Processing (Vol. 1 & 2)*. Orlando, Florida: Academic Press.
- Rumelhart, D E, & McClelland, J L (1986). *Parallel distributed processing: explorations in the microstructures of cognition (1 ed. Vol. 1, Foundations)*. Cambridge, MA: MIT press.
- Russel, S, & Norvig, P (1995). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.
- Saffran, J R, Aslin, R N, & Newport, E L (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Sag, I A, Wasow, T, & Bender, E M (2003). *Syntactic theory: a formal introduction (2nd ed. Vol. 152)*. Stanford, CA: Center for the Study of Language and Information.
- Savage, J E (1998). *Models of computation: exploring the power of computation*. Reading, MA: Addison-Wesley.
- Scannell, J W, & Young, M P (1999). Neuronal population activity and functional imaging. *Proc Biol Sci*, 266(1422), 875-881.
- Schacter, Daniel L, & Tulving, Endel (Eds.) (1994a). *Memory Systems*. Cambridge, Ma: Bradford book.
- Schacter, DL, & Tulving, E (1994b). What are the memory systems of 1994? In E Tulving & DL Schacter (Eds.), *Memory Systems 1994 (pp. 1-38)*. Cambridge, MA: MIT Press.

- Schendan, H E, Searl, M M, Melrose, R J, & Stern, C E (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, 37, 1013-1025.
- Schmidtke, V, & Heuer, H (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research*, 60(1-2), 53-71.
- Scholz, B C, & Pullum, G K (2002). Searching for arguments to support linguistic nativism. *Ling. Rev.*, 19, 185-223.
- Seger, Carol Augart (1994). Implicit learning. *Psychological bulletin*, 115, 163-196.
- Seger, Carol Augart (1998). Multiple forms of implicit learning. In Michael A Stadler & Peter A Frensch (Eds.), *Handbook of implicit learning* (pp. 295-320).
- Seger, Carol Augart, & Cincotta, C M (2005). The roles of the caudate nucleus in human classification learning. *J Neurosci*, 25(11), 2941-2951.
- Seidenberg, M S, MacDonald, M C, & Saffran, J R (2002). Does grammar start where statistics stop? *Science*, 298(5593), 553-554.
- Seidenberg, Mark S (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science*, 275, 1599-1603.
- Senda, M, Ishii, K, Oda, K, Sadato, N, Kawashima, R, Sugiura, M, Kanno, I, Ardekani, B, Minoshima, S, & Tatsumi, I (1998). Influence of ANOVA design and anatomical standardization on statistical mapping for PET activation. *Neuroimage*, 8(3), 283-301.
- Shanks, D R, & Johnstone, T (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *J Exp Psychol Learn Mem Cogn*, 25(6), 1435-1451.
- Shanks, D R, Wilkinson, L, & Channon, S (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *J Exp Psychol Learn Mem Cogn*, 29(2), 248-261.
- Shanks, David R (1994). Human associative learning. In Nicholas John (Ed) Mackintosh (Ed.), *Animal learning and cognition* (pp. 335-374). San Diego, CA: Academic Press.
- Shanks, David R, & St. John, Mark (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367-447.
- Shastri, L, & Ajjanagadde, V (1993). From simple associations to systematic reasoning: A connectionist encoding of rules, variables, and dynamic bindings using temporal synchrony. *Behav. Brain Sci.*, 16, 417-494.
- Shepherd, G M (1997). *The Synaptic Organization of the Brain* (4th ed.). New York: Oxford University Press.
- Shepherd, Gordon M, & Koch, Christof (1998). Introduction to synaptic circuits. In Gordon M Shepherd (Ed.), *The synaptic organization of the brain* (4. uppl. ed., pp. x, 638 s. : ill.). New York, Oxford: Oxford University Press.

- Shieber, S M (1986). *An introduction to unification-based approaches to grammar* (Vol. 4). Stanford, CA: CSLI, Stanford university.
- Shinohara, M, Dollinger, B, Brown, G, Rapoport, S, & Sokoloff, L (1979). Cerebral glucose utilization: local changes during and after recovery from spreading cortical depression. *Science*, 203(4376), 188-190.
- Shinohara, T (1994). Rich classes inferable from positive data: Length-bounded elementary formal systems. *Information and Computation*, 108, 175-186.
- Shinohara, T, & Arimura, H (2000). Inductive inference of unbound unions of pattern languages from positive data. *Theor. Comp. Sci.*, 241, 191-209.
- Siegelmann, H (1999). *Neural Networks and Analog Computation: Beyond the Turing Limit*. Birkhäuser.
- Siesjö, Bo K (1978). *Brain energy metabolism*. New York: John Wiley & Sons.
- Simons, J S, & Spiers, H J (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.*, 4, 637-648.
- Skosnik, P D, Mirza, F, Gitelman, D R, Parrish, T B, Mesulam, M M, & Reber, Paul (2002). Neural correlates of artificial grammar learning. *Neuroimage*, 17(3), 1306-1314.
- Smith, E E, & Jonides, J (1998). Neuroimaging analyses of human working memory. *Proc Natl Acad Sci U S A*, 95(20), 12061-12068.
- Smith, E E, & Jonides, J (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657-1661.
- Smith, J G, & McDowall, J (2004). Impaired higher order implicit sequence learning on the verbal version of the serial reaction time task in patients with Parkinson's disease. *Neuropsychol*, 18, 679-691.
- Smolensky, P (1988). On the proper treatment of connectionism. *Behav. Brain Sci.*, 11, 1-74.
- Squire, L R (1992). *Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans* [published erratum appears in *Psychol Rev* 1992 Jul;99(3):582]. *Psychol Rev*, 99(2), 195-231.
- Squire, L R, Knowlton, B, & Musen, G (1993). The structure and organization of memory. *Annu Rev Psychol*, 44, 453-495.
- Squire, L R, & Zola, S M (1996). Structure and function of declarative and nondeclarative memory systems. *Proc Natl Acad Sci U S A*, 93(24), 13515-13522.
- Stadler, Michael A, & Frensch, Peter A (Eds.) (1998). *Handbook of Implicit Learning*. Thousand Oaks, CA: Sage Publications.
- Stillings, Neil A, Weisler, Steven E, Chase, Christopher H, Feinstein, Mark H, Garfield, Jay L, & Rissland, Edwina L (1995). *Cognitive science: An introduction* (2. ed ed.). Cambridge, Mass: MIT Press.

- Stowe, L A, Paans, A M, Wijers, A A, & Zwarts, F (2004). Activations of "motor" and other non-language structures during sentence comprehension. *Brain Lang*, 89(2), 290-299.
- Strange, B A, Fletcher, P C, Henson, R N, Friston, Karl, & Dolan, Raymond (1999). Segregating the functions of human hippocampus. *Proc Natl Acad Sci U S A*, 96(7), 4034-4039.
- Strange, B A, Henson, R N A, Friston, Karl, & Dolan, Raymond (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, 11, 1040-1046.
- Stromswold, K, Caplan, D, Alpert, N, & Rauch, S (1996). Localization of syntactic comprehension by positron emission tomography. *Brain Lang.*, 52, 452-473.
- Stuss, D T, & Knight, R T (Eds.) (2002). *Principles of Frontal Lobe Function*. Oxford, UK: Oxford University Press.
- Sutton, R S, & Barto, A G (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Tanenbaum, A S (1990). *Structured Computer Organization (3rd ed ed.)*. Englewood Cliffs, NJ: Prentice-Hall.
- Taylor, R G, & Taylor, S (1997). *Models of Computation and Formal Languages*. Oxford, UK: Oxford University Press.
- Tillmann, B, & McAdams, S (2004). Implicit learning in auditory modality: statistics and acoustics. *Revue Neuropsychol*, 14, 147-167.
- Torriero, S, Oliveri, M, Koch, G, Caltagirone, C, & Petrosini, L (2004). Interference of left and right cerebellar rTMS with procedural learning. *J Cogn Neurosci*, 16, 1605-1611.
- Trehub, S E (2003). The developmental origins of musicality. *Nat Neurosci*, 6(7), 669-673.
- Tubau, E, & Lopez-Moliner, J (2004). Spatial interference and response control in sequence learning: The role of explicit knowledge. *Psychol Res*, 68, 55-63.
- Tulving, E (1995). Organization of memory: Quo vadis? In MS Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 839-847). Cambridge, MA: MIT Press.
- Tulving, E, & Schacter, DL (Eds.) (1994). *Memory Systems 1994: The MIT Press*.
- Uddén, Julia, Araújo, Susana, Ingvar, Martin, Hagoort, Peter, & Petersson, Karl Magnus (In submission). Implicit acquisition of recursive non-regular grammars: quantifying the role of grammar classes and violation types.
- Uddén, Julia, Folia, Vasiliki, Forkstam, Christian, Ingvar, Martin, Fernandez, Guillen, Overeem, Sebastiaan, Elswijk, Gijs van, Hagoort, Peter, & Petersson, Karl Magnus (2008). The inferior frontal cortex in artificial syntax processing: An rTMS study. *Brain Res*, 11, 69-78.
- Ullman, M T (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231-270.
- Vapnik, V (1998). *Statistical Learning Theory*. New York: Wiley and Sons.

- Vihman, Marilyn M, & DePaolis, Rory A (2000). *Role of Mimesis in Infant Language*. In Knight & Studdert-Kennedy & Hurford (Eds.), *The evolutionary emergence of language*.
- Voermans, N C, Petersson, K M, Daudey, L, Weber, B, van Spaendonck, K P, Kremer, H P H, & Fernández, G (2004). *Interaction between the human hippocampus and caudate nucleus during route recognition*. *Neuron*, 43, 427-435.
- Vokey, J R, & Brooks, L R (1992). *Saliency of item knowledge in learning artificial grammar*. *Journal of experimental psychology. Learning, memory, and cognition*, 18, 328-344.
- von Neumann, John (1957). *The computer and the brain*. New Haven: Yale University Press.
- Vonsattel, J P, Myers, R H, Stevens, T J, Ferrante, R J, Bird, E D, & Richardson, E P, Jr (1985). *Neuropathological classification of Huntington's disease*. *J Neuropathol Exp Neurol*, 44(6), 559-577.
- Vosse, T, & Kempen, G (2000). *Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar*. *Cognition*, 75(2), 105-143.
- Walker, M P, Brakefield, T, Morgan, A, Hobson, J A, & Stickgold, R (2002). *Practice with sleep makes perfect: Sleep-dependent motor skill learning*. *Neuron*, 35, 205-211.
- Walsh, Vincent, & Pascual-Leone, Alvaro (2003). *Transcranial magnetic stimulation : A neurochronometrics of mind*. Cambridge, Mass.: MIT Press.
- Wasow, T (1989). *Grammatical theory*. In M Posner (Ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- Wells, Andrew (2005). *Rethinking cognitive computation: Turing and the science of the mind*. Basingstoke: Palgrave Macmillan.
- Wernicke, C (1874). *Der Aphasische Symptomen Komplex*. Breslau: Cohn and Weigart.
- Wiener, N (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. New York: Technology Press.
- Wilkinson, L, & Shanks, D R (2004). *Intentional control and implicit sequence learning*. *J Exp Psychol Learn Mem Cogn*, 30, 354-369.
- Wilson, Robert A, & Keil, Frank C (Eds.) (1999). *The MIT encyclopedia of the cognitive sciences*. Cambridge, Massachusetts; London England: MIT Press.
- Wise, S P, Murray, E A, & Gerfen, C R (1996). *The frontal cortex-basal ganglia system in primates*. *Crit Rev Neurobiol*, 10(3-4), 317-356.
- Wittgenstein, Ludwig (1953). *Filosofiska undersökningar*.
- Wood, J N, & Grafman, J (2003). *Human prefrontal cortex: Processing and representational perspectives*. *Nat. Rev. Neurosci.*, 4, 139-147.

- Worsley, K J, Marrett, S, Neelin, P, Vandal, A C, Friston, Karl, & Evans, A C (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4, 58-73.
- Young, M P, Hilgetag, C-C, & Scannell, J W (2000). On imputing function to structure from behavioral effects of brain lesions. *Phil. Trans. R. Soc. Lond. B*, 355, 147-161.
- Zajonc, Robert (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 9, 1-27.
- Zarahn, Eric, Aguirre, Geoffrey, & D'Esposito, Mark (1997). Empirical analyses of BOLD fMRI statistics I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage*, 5, 179-197.
- Zizak, Diane, & Reber, Arthur (2004). Implicit preferences: the role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition: An International Journal*, 13, 336-362.
- Zurif, E (1990). *Language and the brain*. In D N Osherson & H Lasnik (Eds.), *Language: An Invitation to Cognitive Science, Volume 1*. Cambridge, MA: MIT Press.
- Zurif, E (1998). The neurological organization of some aspects of sentence comprehension. *J. Psycholing. Res.*, 27, 181-190.

CHAPTER 8 — INCLUDED STUDIES