Mirko Hanke

Strategies and mechanisms for language production in noise

Self-published/im Selbstverlag
Erlensee 2014
Strategies and mechanisms for language production in noise

Von der
Carl von Ossietzky Universität Oldenburg,
Fakultät III Sprach- und Kulturwissenschaften,
zur Erlangung des Grades eines
Doktors der Philosophie (Dr. phil.)
genehmigte Dissertation von

Mirko Hanke
geboren am 3. November 1979 in Hanau
Referentin: Prof. Dr. Cornelia Hamann
Koreferent: Prof. Dr. Robert Hartsuiker (Universität Gent, Belgien)
Koreferentin: Prof. Dr. Esther Ruigendijk

Tag der Disputation: 27.6.2014
Contents

Acknowledgements ix

1. Introduction 1
   1.1. Motivation ........................................... 1
   1.2. Research question .................................... 3
   1.3. Approach ............................................. 4
   1.4. Thesis outline ....................................... 5

I. Theory 9

2. Sentence structure complexity 11
   2.1. Introduction .......................................... 11
   2.2. Some structural problems ......................... 13
       2.2.1. Active and passive ............................... 13
       2.2.2. Relative clauses ................................. 17
       2.2.3. Agreement ......................................... 22
   2.3. Linguistic complexity .............................. 26
       2.3.1. What is complexity? .............................. 26
       2.3.2. Formalising language .......................... 27
       2.3.3. Linguistic metrics ............................... 30
       2.3.4. Grammar theory and the goals of investigation . 33
   2.4. Summary ............................................... 36

3. Sentence processing-for-production 39
   3.1. Introduction ........................................... 39
   3.2. Levelt’s language production architecture .......... 41
       3.2.1. Incrementality, seriality, and automaticity .... 42
       3.2.2. Monitoring ......................................... 44
       3.2.3. Embedded grammar ................................ 45
### Contents

3.3. Linguistic and psycholinguistic modelling ................................. 46
  3.3.1. Grammar theory and processing mechanisms ........................ 47
  3.3.2. Marr’s Three-Tier-Hypothesis ........................................ 50
  3.3.3. Competence and performance ........................................... 51

3.4. Processing of different phenomena ........................................ 53
  3.4.1. Active and passive ....................................................... 55
  3.4.2. Relative clauses ........................................................... 58
  3.4.3. Agreement ................................................................. 61

3.5. Summary ............................................................................ 69

4. Psychological aspects of speaking in noise ................................. 71
  4.1. Introduction ........................................................................ 71
  4.2. Individual differences in processing-for-production ................. 72
    4.2.1. Effects of hearing impairment ........................................ 72
    4.2.2. Effects of ageing ........................................................... 73
  4.3. Moment-to-moment differences in performance ....................... 75
    4.3.1. Effects of noise on language comprehension .................... 75
    4.3.2. Effects of noise on speaking ......................................... 78
    4.3.3. The irrelevant sound effect (ISE) ...................................... 80
  4.4. Cognitive resources and interference ..................................... 80
    4.4.1. Dual-task performance ................................................. 81
    4.4.2. Working memory .......................................................... 84
    4.4.3. The changing-state hypothesis ....................................... 88
  4.5. Summary ............................................................................ 90

II. Data ...................................................................................... 93

5. Study 1: Near-spontaneous speech production ........................... 95
  5.1. Introduction ........................................................................ 95
  5.2. Experiment 1: Interview and picture story description ............ 95
    5.2.1. Spontaneous production ................................................. 95
    5.2.2. Measures for sentence structure complexity .................... 96
    5.2.3. Method ........................................................................ 98
    5.2.4. Results ....................................................................... 103
  5.3. Discussion .......................................................................... 106
    5.3.1. Reading span ................................................................ 106
8.2. Summary of the empirical results .................................. 166
  8.2.1. Effects on structural complexity ......................... 166
  8.2.2. Effects on the amount of errors ....................... 167
  8.2.3. Effects on utterance latency and speech rate ....... 167
8.3. Interpretation ..................................................... 168
  8.3.1. Strategy ....................................................... 168
  8.3.2. Working memory mechanisms .................... 169
  8.3.3. The role of the monitor ............................. 174

9. Conclusion and outlook ............................................. 177

A. Stimulus material study 1 ........................................... 179
  A.1. Interview questions: Holidays ............................ 179
  A.2. Interview questions: Travel ................................ 179
  A.3. Short picture story: Well intended ...................... 181
  A.4. Short picture story: Peacemaker ....................... 182

B. Stimulus material study 2 ......................................... 183
  B.1. Exp. 2.1: Active/passive sentence priming ............ 183
  B.2. Exp. 2.2: Relative clause elicitation .................. 193

C. Stimulus material study 3 .......................................... 203

Bibliography .............................................................. 205

Zusammenfassung ....................................................... 239
Acknowledgements

Writing a doctoral thesis is never a lonely enterprise—sometimes this is easy to forget, especially when you are in the midst of the process. Quite a number of people have crossed the winding path that led to this finished book and I would like to give some credit for the advice, criticism and support I received. To all whom I might have unfairly forgotten here, my apologies.

First and foremost, my sincere thanks goes to my advisor Cornelia Hammann, for giving me the opportunity to work with her, for her advice and encouragement, and for her loyal support throughout my entire time in Oldenburg. I am convinced that our challenging discussions on theoretical linguistic matters broadened my horizon considerably and made this thesis better. I am also deeply indebted to my co-advisor and referee Esther Ruigendijk. Not only provided she my colleagues and me with an indispensable and finely equipped laboratory. She also advised me with many methodological and psycholinguistic questions, and was a keen and always encouraging critic of my writing. My external referee, Robert Hartsuiker of Ghent University I have to thank for his inspiring comments and always constructive criticism. Birger Kollmeier I thank for his critical thoughts that helped to sharpen the goals of my research, as well as for his generous help in times of need.

Next I would like to thank my colleagues, first of all Rebecca Carroll, with whom I shared an office for most of my time in Oldenburg. Without the many hours of discussion across our desks and in front of our whiteboard this thesis would never have materialised. Only in retrospect I realize how much your drive to do research inspired me, please keep that going. Your open-minded and cordial attitude made arriving in a new town easy, and the time there would have been far less fun without you. In addition I owe many thanks to Sara Jonkers, who shared the office with me for about a year and who was an invaluable discussion partner. Also, a big thank you for inspiring the love for Belgian delicacies. Then, I want to express my
Acknowledgements

sincere gratitude to my colleagues from the AULIN research project, Tom Brand, Verena Uslar and Dorothea Wendt: Thanks for the support and numerous discussions in and outside of work group meetings, and thanks for the experience that doing interdisciplinary work is demanding, but also a very rewarding enterprise. Further I want to give a big and well-deserved thank you to Hendrikje Ziemann, as well as the project’s student assistants Sarah Bauer, Jonas Fortmann and Heike Viohl for their indispensable help in collecting and crunching data. Many thanks also to the technical and reception desk staff of the Oldenburg House of Hearing audiological research centre, in particular Niklas Grunewald, for technical and administrative support. From the linguist crowd at our faculty I would like to give my special thanks to Kristian Berg, Franziska Buchmann and Tina Morcinek for the hours of discussions about grammar, linguistics, life and everything.

Further thanks to: Franziska, for acquainting me with Twin Peaks, and for baking scrumptious blueberry pie; to Gesa, Ina and Sara for being my faithful Doppelkopf-buddies; to Tina and Götz for remaining friends with me after all that happened; to Hans Scharoun for being such a great architect.

Last but not least, I thank my family, Cornelia Ruth-Hanke, Reinhard Hanke and Daniel Hanke for many, many years of unfailing love and support during my long, long education. Finally, thanks to my husband José Ramón Alvarado, whose unconditional affection for me still makes my heart tingle, and whom I love beyond words.
1. Introduction

1.1. Motivation

Our ability to speak is one of the most fascinating human capacities, perhaps one of the core characteristics of humanity. Speaking appears to most healthy adults as a very natural thing to do; we do it all the time (some more so than others), and seemingly we produce language without much conscious effort.

We produce (and comprehend) language under various different circumstances: Imagine the typical psychological or (psycho-)linguistic department and its laboratory setting, with carpet on the floor to silence steps, hushed voices, and sometimes even sound-attenuated rooms, in which test subjects are made to speak or listen to language. Now imagine a large city, with cars, buses, trams, people shouting, pedestrians talking over the noise to others on the phone. Admittedly, these two situations form a rather extreme opposition, but it should be intuitively clear from these imagined settings that the presence of external noise can impose the feeling of greater effort in both speaking and comprehending. There are some situations in which speakers will intuitively experience speaking as more cumbersome or effortful than usual, for instance when many people talk at the same time or when other background noise is present.

In audiological and psychological literature, the negative influence of distracting noise on spoken language perception and comprehension has been studied for several years. A perceived perception or comprehension difficulty in the presence of other people speaking at the same time is usually referred to as “cocktail party effect”, see for instance Bronkhorst (2000) for a review from an audiological perspective. In audiology, the effects of noise on speech perception ability have been in the focus of interest, because ‘noisy’ real-world settings can pose considerable difficulty to people with hearing impairments. Some methods for screening hearing acuity thus try to simulate the cocktail party effect in an experimental setting, in order
1. Introduction

to yield an estimation of the subject’s speech perception performance in everyday situations (e.g. Hagerman, 1982; Kollmeier & Wesselkamp, 1997; Plomp & Mimpen, 1979).

The possible influence of difficult communication settings on language comprehension has also come increasingly to the attention of psycholinguists in recent years. Pichora-Fuller, Schneider and Daneman (1995) for instance have shown that noise affects comprehension and memory retention of spoken language. Their results further indicate that the negative effects of noise interact with other adversary factors like hearing impairment and cognitive ageing. What is more, comprehension effort in noise appears to be additionally modulated by the syntactic difficulty or complexity of language presented to listeners (McCoy et al., 2005; Wingfield, McCoy, Peelle, Tun & Cox, 2006).

The can’t hear myself think feeling experienced in a noisy street or typical cocktail party setting might also have an influence on how we speak. For instance, Ladefoged (1967) reports an “informal experiment” on speech production, during which auditory feedback on speaker’s own speech was removed by presenting them with very loud noise over headphones. The resulting speech is characterised by Ladefoged as “disorganised”, but still intelligible (p. 163).

However, despite Ladefoged’s early interest in the problem and compared to the amount of data available on language comprehension under acoustically difficult circumstances, experimental evidence on language production in noise is rather limited in scope. A considerable body of work on speaking in noise has focused on prosodic or articulatory effects of adverse acoustics, in research on the so-called Lombard reflex or effect: Under noise, speakers perceive an increase in vocal effort. In addition, language produced in noise is perceived with greater loudness. Finally, spectral changes have been reported. Beyond this line of work it remains to a large extent an open question how acoustic difficulties like the presence of noise can influence language production at other linguistic levels, including the level of syntactic structure.

Furthermore, while empirical data on syntactic aspects of language production under adverse conditions like noisy environments is scarce to begin with, controlled experimental studies on this topic are even scarcer. Generally, sentence production is a complex process spanning multiple stages of linguistic processing which are not, or rather indirectly available to ob-
servation. Studies in theoretical linguistics, corpus-based linguistic studies, and research on language acquisition have been dealing with sentence generation from an off-line perspective, by analysing patterns in the result or product of the speaking process. Experimental evidence about the on-line or real-time characteristics of sentence production is harder to come by and therefore comparatively rare.¹

A rather solitary exception to the lack of experimental studies on speaking in noise is work done by Kemper, Herman and Lian in 2003. The authors found an interaction between the effects of noise and ageing on syntactic complexity of spoken utterances, measured by counting the number of embedded sentences and other structures in order to calculate two different complexity scores. The notable work by Kemper and colleagues leaves some aspects unaddressed, however: for instance the question whether there are other linguistic properties of sentence structure that can result in the perception of greater complexity or difficulty, and it remains open how their observations would translate into a processing model for language production. Finally, to my current knowledge, none of the previous research has been dedicated to timing aspects of sentence structure generation in noise.

1.2. Research question

With the observational studies and experiments reported here, I will complement the existing research on the influence of noise on language production at a structural level, taking Kemper et al. (2003) as a starting point. The research for this thesis was aimed at providing more detailed insight into language production processes. As a part of human cognitive capacities, these processes do their work under noisy, far from perfect conditions.

Crucially, the language production system’s processing stage responsible for generating sentence structure has been argued to operate in a highly automatic fashion (Levelt, 1989, p. 22), and relatively independent of (or isolated from) other aspects of cognition:

¹Electrophysiological measures for instance, useful to assess the time course of linguistic processes in the brain during comprehension, can be used only to a very limited extent for research on language production because the muscular activity during speaking creates severe distortions of the measurement.
1. Introduction

A tacit assumption underlying much of the empirical research on language processing is that the linguistic components are largely independent of other cognitive processes (A. S. Meyer, Wheeldon & Krott, 2007: p. x)

Given the intuitive experience of greater effort when speaking in noise and the previous findings on this topic, a number of interesting questions follow from the assumption about the automaticity of language production. The main question I tackled experimentally was how the processes involved in creating syntactic structure representations of varying complexity are affected by different, momentary communication settings or performance conditions, analogous to earlier investigations on language comprehension. More particularly, I was interested to explore the effects of background noise on speaking, as one particularly common example of adverse conditions in everyday communication. With respect to existing models of language processing-for-production, I pursued the question how the additional necessity to ‘filter’ background noise can influence speaking, and where interference might arise.

1.3. Approach

In the research reported here I adhered to theoretical concepts and methods from psycholinguistics. More specifically, I followed the tradition of the information processing approach from cognitive psychology (e.g., Harley, 2001; Horst, 2011), which metaphorically treats the mind as a computer processing data. Within the context of this view, researchers aim to model pathways, timing, and boundary conditions of data or information flow during cognitive activity such as language processing. A general goal for psycholinguistic research, based on programmatic desiderata issued by for instance Jackendoff (2002, 2007) or Dell, Oppenheim and Kittredge (2008), is to understand sentence processing mechanisms within the larger context of human cognition. Accordingly, it has formed an increasing trend to integrate knowledge about neighbouring domains of cognition in psycholinguistic work (Alario, Costa, Ferreira & Pickering, 2006). The work presented here strives for interdisciplinarity by aiming to identify interfaces between theories from related psychological domains.
Given the scarcity of previous research on sentence generation in noise, further empirical exploration of the phenomenon seemed necessary as a first step. For this project I carried out quantitative studies, using three observational and experimental paradigms to capture different aspects of the problem. In order to set necessary limits to the scope of the research program I focused mostly on the case of healthy adults speaking under adverse conditions, while keeping difficulties related to language acquisition out of immediate consideration. This way, the problem could be clearly demarcated as a performance or processing issue, with no immediate repercussions on the competence of the speaker.

The research reported here goes beyond previous work on 'higher order', structural aspects during speaking, in that the acoustic setting during data collection was kept under narrow control. The studies used speech-free noise signals to model noisy environments. This way, possible effects of linguistic content in the noise could be eliminated. In addition, the stimulus material used for the experimental studies was explicitly designed to show different, controlled levels of structural complexity, and it was thoroughly pre-tested. Finally, I included data related to the time-course of producing complex syntactic structure in noise in my analysis, which to my current knowledge has been without precedent so far.

1.4. Thesis outline

This thesis is divided into three parts, dedicated to the theoretical background, empirical observations, and discussion and interpretation respectively. At the beginning of part one, in Chapter 2, I will discuss the question how to operationalise the quantification of syntactic complexity. A number of different sentence structure types will serve as examples to show how syntactic theories have been formalising structural complexity. I will extend the discussion to more epistemological questions about the subject of investigation of linguistic theory, in order to show that a formal linguistic model and a processing theory necessarily need to complement each other to provide an explanation of complexity phenomena. Chapter 3 will explicate the theoretical foundations for a sentence production model, with special focus on the generation of sentence structure considered complex or difficult from the viewpoint of syntactic theory. I will continue to discuss
the relationship between linguistic theory and processing models, this time from a more psycholinguistic perspective. The chapter is concluded with a review of experimental evidence on the production and comprehension of different types of complex sentence structure. Chapter 4 covers the psycholinguistic and more generally psychological question how the mental processing of different kinds of 'data', in this case noise and sentence structure, can interact. In the discussion I will cover recent insights by other authors about some fundamental theoretical concepts or constructs like working memory, processing resources and dual-task interference. With these cornerstones I will then lay the foundation for a preliminary conceptual model of noise interference during speaking and I will point out some more general implications for interference between different cognitive 'modules' or functions in general.

Part two presents the empirical building blocks for this work. The research programme was meant to ‘zoom in’, both methodologically as well as in terms of the architecture of the language production system, in order to get a view from different angles. Beginning with (near-)spontaneous, undirected speech in the study reported in Chapter 5, I looked for effects of noise that might be apparent in a corpus of picture story descriptions and interview answers. Chapter 6 will report a more constrained study, in which I elicited sentences of varying complexity in a picture description task. The rationale for this study was to take a more detailed look at structures already known to pose difficulty in processing. In addition, the factor hearing impairment was added to check for possible “effortfulness” effects that have been observed in studies on language comprehension (e.g., Rabbitt, 1968; Wingfield, Peelle & Grossman, 2003). The final experiment reported in Chapter 7 dealt with morphosyntactic aspects of language production and hence more fine-grained observations about the effect of noise on performance. In order to find out more about the possible location of interaction effects, the experiment tackled agreement production, since this syntactic process crucially involves both syntactic as well as (morphophonological) form aspects.

The general discussion in Chapter 8 will put empirical observations from

---

²The following chapters will include some discussion of relevant processing data, and where appropriate this will also include data on processing-for-comprehension—this is mostly because only little or next to no experimental evidence on processing-for-production is available.
part two and theoretical concepts as covered in part one in relation. I will present theoretical implications of the experiments, and discuss the methodological challenges I faced. Finally, Chapter 9 will summarise the key findings in a conclusion, and point to various open questions that warrant further investigation.
Part I.

Theory
2. Sentence structure complexity

2.1. Introduction

While the focus of the work presented in this thesis is experimental, the notions of complexity that formed the basis for the construction of stimulus material need to be explicated. The notion of structural or syntactic complexity forms a recurrent theme in linguistic as well as psycholinguistic literature. In his typological approach to the notion of ‘markedness’ in grammar, Givón (1991) takes structural complexity to be one of the hallmarks of marked structure.¹ An interesting fact that remains to be explained, according to Givón, is that structural complexity and other criteria like frequency distribution tend to be correlated (ibid., p. 337). Many researchers in psycholinguistics have tackled that question and went beyond Givón’s assumption of a mere correlation between frequency and structural complexity by testing to what extent the latter can serve as an explanation of the former (or vice versa). In fact, in many areas of psycholinguistic research the concept of structural complexity is invoked as a factor meant to explain disparate phenomena of linguistic behaviour, like typical and atypical acquisition order for particular structures, the frequency of structure types in corpora, processing difficulties with particular structures in children or pathological populations, or on-line measures for processing of various structure types. Research which focused on the on-line processing of different types of structure has in some cases corroborated the results

¹See however Haspelmath (2006) who argues that the term ‘markedness’ is often used to describe apparently ‘complex’ or ‘unusual’ structure without sufficient theoretical constraints, leading to problems of circularity, and an undesirable polysemy in the term’s usage. He suggests that for most uses the term can be safely replaced by more substantive concepts, such as frequency or phonetic length. The definition by Givón (1991) is already a first step in this direction, as it separates different factors that have been influencing linguists’ conceptions of ‘markedness’, and among those is structural complexity.
from observational studies, while in other cases contradictory results about the role of structural complexity have been found.

In order to come up with theoretical grammatical and processing accounts for linguistic patterns, grammarians and psycholinguists alike have described sentence structure driven by the goal to identify factors that are seen as indicative of higher complexity. Yet why is complexity of sentence structure interesting to psycholinguists? I assume the view that the human brain is an information processing system, a ‘computer’, as commonly maintained in many areas of cognitive science (cf. e. g. Jackendoff, 2002; Levelt, 1974/2008). This view entails that cognition, including language processing for both production as well as comprehension, is a form of (mental) computation.² Analysing the complexity of linguistic data structure therefore allows us a view at what the brain is able to compute (Jackendoff, 2002: 27; Frank, 2004)—according to Jackendoff, language presents an interesting test case for investigating the processing of (‘complex’) mental structure.

However, despite its widespread use as both an observed variable and explanatory tool in linguistics and psycholinguistics, structural complexity of sentences has remained a rather elusive concept. To this day there is no single, unanimously accepted definition available for what constitutes structural complexity, and the operational definitions used for different studies vary greatly (cf. Kusters, 2008).

In this chapter I will discuss possible substantiations of the term ‘complexity’ both in a rather general sense as well as more particularly with respect to syntactic complexity. At the beginning of this chapter I will provide linguistic descriptions of different structural phenomena. In the empirical studies carried out for this thesis I have used different sentence structure types and I will discuss representational complexity issues related to these structures as well as how the phenomena are treated in different theoretical frameworks. The second half of Chapter 2 will provide a more abstract discussion of how complexity in general might be treated formally, and in Section 2.3 of this chapter I will briefly outline several attempts from the (psycho-)linguistic literature to formally capture structural complexity as a characteristic of a sentence structure representation. These considerations will be complemented in Chapter 3 with a treatment of processing issues and experimental results in research about processing difficulty.

²See however Searle (1980) for a fundamental philosophical criticism of this view.
2.2. Some structural problems

Three types of structure have been chosen as the primary objects of study for the project reported here. They form exemplars of phenomena that have traditionally been linked with structural complexity in the literature: passive sentences, sentences with subject-extracted relative clauses in which the relative pronoun functions as the object of the embedded clause (‘object relative clauses’), and agreement relations in sentences with an embedded constituent.

2.2.1. Active and passive

The \textit{diathesis} or alternation between active and passive voice as in sentences (1), (2) and (3) has been a subject of grammatical research for decades.

(1) Active: Der nette Clown umarmt den frechen Kasper.
\text{The}_{\text{nom}} \text{nice clown} \text{hug}_{3\text{sg}} \text{the}_{\text{acc}\text{; masc}} \text{cheeky buffoon}.
The nice clown is hugging the cheeky buffoon.

(2) Agentless passive: Der freche Kasper wird umarmt.
\text{The}_{\text{nom}} \text{cheeky buffoon} \text{become}_{3\text{sg}} \text{hug}_{\text{ptcp}}.
The cheeky buffoon is being hugged.

(3) Passive: Der freche Kasper wird vom netten Clown umarmt.
\text{The}_{\text{nom}} \text{cheeky buffoon} \text{become}_{3\text{sg}} \text{by}_{\text{dat}} \text{the}_{\text{acc\text{; masc}}} \text{nice clown} \text{hug}_{\text{ptcp}}.
The cheeky buffoon is being hugged by the nice clown.

Many languages exhibit passive, or passive-like constructions, in a variety of forms, see for instance Kazenin (2001) or Haspelmath (1990) for an overview. A characteristic of passive structures in languages like German or English is that an element which serves the object function in an active sentence will be assigned the subject function in the corresponding passive sentence, while the subject of the active sentence is either omitted or introduced with a prepositional phrase, compare examples (2) and (3). Morphologically, both synthetic as well as analytical marking can be observed across languages. In English and German we see a combination of synthetic and analytical marking with an auxiliary (a form of \textit{to be} in English, \textit{werden}
2. Sentence structure complexity

[to become] in German, in some cases also with bekommen [to get]) and a particular morphological form of the verb (past participle). Languages differ with respect to the possibility whether the subject of a corresponding active sentence can be introduced, and with respect to the way this is done. In German the demoted or ‘underlying’ subject argument can be mentioned as part of a phrase headed either by an instrumental (durch) or a locative (von) preposition, cf. example 3.

Kazenin (2001) classifies the passive functionally as part of a group of de-transitivization phenomena (p. 899), which also includes anticausative and medio-passive structures. According to the author, it is still debated among typologists whether the phenomena traditionally classified as passive can be reduced to a particular function it expresses in most, if not all languages. Proposals center around the promotion of non-Agent role bearers and/or the demotion of the Agent. That is, one basic generalisation about passive constructions is either that they allow to express the Theme or Patient of an event (or other non-Agent role bearers comparatively low in the thematic hierarchy; cf. Dowty, 1991; Jackendoff, 1987) in a ‘prominent’³ position, with optional omission of the Agent. Or, we might say alternatively that the passive allows the Agent role to be left unexpressed (Givón, 1981; Haspelmath, 1990). Givón (1981) and Haspelmath (1990) summarise this with their generalisation that the event semantics of a proposition which is expressed by a passive sentence changes (compare the recent rediscovery of this insight by Gehrke & Grillo, 2009). This change has an effect on the morphological realisation, cf. for instance the definition given by Haspelmath (1990).⁴

Different grammatical frameworks represent the voice diathesis with different formal tools, for instance

³Typically the canonical subject position towards the left edge of the sentence in languages like German or English.
⁴The definition of ‘passive’ proposed by Haspelmath (1990) requires

1. the passive marking to be morphologically derived from the respective active morphology
2. arguments bearing non-Agent or non-Actor roles to be mapped to a higher grammatical function (“promoted”) and/or
3. the Agent or Actor role-bearer of the corresponding active sentence to be dropped or mapped to a lower (object) grammatical function (“demoted”).
• as one of different transformational/movement-based accounts from the realm of mainstream generative grammar (MGG; e. g. Baker, Johnson & Roberts, 1989; Collins, 2005; Gehrke & Grillo, 2009; Jeggli, 1986).

• with recourse to lexical rules that change the verb’s argument structure, as for instance assumed by lexical-functional grammar (LFG; Bresnan, 1982) and head-driven phrase structure grammar (HPSG; e. g. S. Müller, 2008; Sag, Wasow & Bender, 2003)

• with construction or schema-based approaches, as for instance in variants of construction grammar (CxG; e. g. Goldberg, 1995, 2003), or ‘Simple syntax’, (Culicover, 2009; Culicover & Jackendoff, 2005)

Historically, the intuitively apparent semantic similarity between a sentence in active and in passive voice has been captured in Chomsky’s 1965 ‘Aspects’ model by one of the original transformations. The model postulated that two sentences expressing the same proposition might originate with identical deep structure representation, but are mapped in different ways to surface realisations. Non-transformational formalisms like LFG or HPSG apply lexical rules that change the argument structure and subcategorisation properties of passivised verbs in the lexicon, and this way the grammar licenses the mapping of the Patient (or another non-Agent role bearer) to the subject position; see for instance Sag et al. (2003) or S. Müller (2008, 2010). More recent MGG accounts within the Government and Binding (GB) framework assume for instance that the passive morphology ‘triggers’ the generalised transformation movement, and that the internal argument is ‘smuggled’ along with a participle across a movement barrier (Collins, 2005). The recent analysis of passive structure suggested by Gehrke and Grillo (2009) is based on the minimalist program (MP; Chomsky, 1995) framework and instead suggests that syntacticised parts of event structure are moved to a functional projection for event time.

Despite the superficial differences with non-transformational accounts in terms of the formal representation, many accounts of the passive that rely on transformations or movement operations thus also hinge on the assumption that argument structure properties and case assignment features of the participle form of a verb differ from those of a verb in active voice. In this respect, transformational and non-transformational accounts
2. Sentence structure complexity

are rather similar, differing mainly in the place where the argument structure changes become apparent (lexicon vs. syntactic derivation; S. Müller, 2010: 383). Somewhat in between lexicalist or non-transformational and transformational accounts falls the proposal of Culicover and Jackendoff (2005) for a ‘Simpler Syntax’ analysis. Their model captures the same intuitions about the relative order of arguments in so-called ‘correspondence’ templates, which mediate between representational tiers for argument/conceptual structure, grammatical function and syntactic categories. Such a correspondence formally expresses for instance the binding of a conceptual structure’s Theme to the highest grammatical function on an intermediate grammatical function tier (see also Culicover, 2009). For the passive, Culicover (2009), and Culicover and Jackendoff (2005) assume ‘crossed’ bindings between ordered elements on the grammatical function, syntactic structure and semantic structure tier. A similar approach based on typed correspondence schemata linking word order patterns and information-structural or semantic function has been suggested in variants of Construction Grammar; see e.g. Goldberg (1995).

Different structural factors have been discussed as leading to relatively higher ‘complexity’ of passive sentences. Common to formal analyses in any framework is that passive sentences in German or English require additional structure or rules to be represented, compared to active sentences (cf. Culicover & Jackendoff, 2005: 205). This is especially obvious in cases where the subject of the active counterpart is realised in an overt prepositional phrase in the passive sentence, because only in those cases it can be reasonably assumed that an equivalent amount of propositional content is expressed (compare examples 2 and 3 on page 129). Another factor is the relative order of argument bearers in the sentence surface structure (‘canonicity’). In formal treatments within the generative mainstream the ‘unusual’, non-canonical word order of passives has been analysed as a case of structures that require additional movement operations, and are in violation of syntactic locality of dependencies (Baker et al., 1989; Chomsky, 1995; Jaeggli, 1986; Rizzi, 2004).³

³Note that in essence, canonicity-based explanations imply the assumption that the active argument order and the mapping of that order to syntactic structure is ‘canonical’, or more ‘basic’ (whatever this might mean) than the order expressed in passive sentence. Drai and Grodzinsky (2006) and Koster (2010) link this question to properties of a verb’s lexicon entry, in particular the argument structure from which the syntactic
One formal construal of syntactic locality is Rizzi’s concept of relativised minimality (RM for short; cf. Rizzi, 1990, 2004). It is formulated as a grammatical principle, according to which an element with a particular set of grammatical features forms a barrier for establishing a dependency across that element. More particularly, the principle is formalised as follows: in a constellation $X << Z << Y$, the establishing of a syntactic dependency between $Y$ and $X$ is blocked or dis-preferred if there is an intervening element $Z$ that is “structurally similar” to $Y$, i.e. bears a similar⁶ set of syntactic features.

The RM principle has been proposed as an explanation for canonicity effects that make passive sentences “dis-preferred” viz. active sentences (cf. Belletti, 2009; Grillo, 2008).⁷ Transformational accounts have argued for higher complexity of passive sentences along the following lines: passive sentences that involve movement of an object over the by-phrase would be potentially subject to violations of locality.

### 2.2.2. Relative clauses

Relativisation of nominal elements appears to be quite widespread a phenomenon in the world’s languages. **Relative clauses** in European languages configuration ‘projects’.⁶

An important aspect of Rizzi’s formalisation is the concept of ‘similarity’. Despite being intuitively rather straightforward, it has proven difficult to formalise mathematically (see Ennis, 2007; Lin, 1998; Suppes, 2008 for discussion). In order to linguistically substantiate the notion of similarity for his definition of RM, Rizzi suggested a typology of feature specification on morphological/lexical elements. The set of feature categories he suggests has been tied in with the minimalist conception of functional categories and feature-driven movement (feature checking; cf. e.g. Radford, 1997), and relates to work from the so-called ‘cartographic’ approach to syntactic representations, which links the syntactic configuration to functional properties of a sentence, similar to so-called topological fields, or functionally specified templates in Construction Grammar (cf. S. Müller, 2010: 91). For the sake of his approach, Rizzi defines two elements to be similar if they contain features from the same category.

Note however that the formulation of RM as a grammatical principle implies that structure which is in violation of that principle is ungrammatical—which is clearly not the case with passive sentences. Therefore, newer accounts from the generative mainstream try to avoid the possibility of RM violations in the analysis of passives, like the ‘smuggling’ analysis by Collins (2005) or the approach based on sub-lexical feature movement suggested by Gehrke and Grillo (2009), see Belletti and Chesi (2011) for discussion. Also see Section 2.3.4.
2. Sentence structure complexity

like English or German contain a proposition that modifies a nominal element or another correlate in a matrix clause. Usually a distinction is made between restrictive and appositive (or ‘free’) relative clauses, depending on the type of relation between the relative clause and the “head” it modifies (cf. examples 4 and 5; see Lehmann, 1995, for a typology and see Bianchi, 2002a, 2002b for an overview over formal treatments).

(4) Der Schüler, der die Hausaufgaben nicht gemacht hat, muss nachsitzen.  
The student who didn’t do his homework, has to follow extra lessons.

(5) Die Sonne, die jeden Morgen im Osten aufgeht, wärmt uns.  
The sun, which rises every morning in the east, is spreading warmth.

According to Lehmann (1995), a functional definition of restrictive relatives is possible based on the central semantic feature of these clauses to act as a constraint on the extension of category expressions. Appositive relative clauses instead add semantic information, without further restricting the relativised element’s extension. Because means of relativisation vary across languages, Lehmann seems to be sceptical about the possibility to match his functional definition with a similarly uniform structural definition; also see Keenan and Comrie (1977) or De Vries (2002) on this issue.

In the following, I will concentrate on a particular type of structural relativisation strategy: the restrictive, externally headed relative clause as it can be found in many European languages including German (cf. example 4). This type of sentence structure allows for a rather strictly controllable syntactic (as well as semantic) variation. Consider the following examples:

(6) Der Hund, der die Katze verfolgt, bellt.  
The dog, who is chasing the cat, is barking.

(7) Der Hund, den die Katze verfolgt, bellt.  
The dog, whom the cat is chasing, is barking.

Examples 6 and 7 are characterised by the fact that the relative clause modifies the subject of the matrix clause.⁸ Of particular interest for the

---

⁸It is possible to construct sentences in which the object of a matrix clause is modified by a relative clause, consider the following examples:
2.2. Some structural problems

studies to be reported here is the respective grammatical role or function the relative pronoun has within the relative clause. In example sentence 6 the relative pronoun bears the subject function, whereas in example 7 it is the object function.⁹ For the remainder of this thesis, I will refer to the first type of sentences as *subject relative clause* (SR) structures, and the latter as *object relative clause* (OR) structures.

Bianchi (2002a, 2002b) provides an extensive review of relativisation from a mainstream generative perspective. According to her review, a grammatical analysis of relative clause structures has to capture the link between the relative pronoun and the canonical position in which a constituent is interpreted as subject or object of a relative clause (a ‘trace’), and the link between the relative clause and the element which it modifies. Grammar formalisms differ in how they capture these links or dependencies, for instance by feature percolation (as in HPSG, cf. for instance S. Müller, 2008; Sag et al., 2003), binding on the conceptual structure level (Simpler Syntax, cf. Culicover, 2009; Culicover & Jackendoff, 2005), or by movement of constituents/formation of dependency chains as in most analyses from the field of mainstream generative grammar (Bianchi, 2002a, 2002b).

(1) Der Hund jagt die Katze, die gerne Mäuse frisst.
   The dog chases the cat, who likes to eat mice.

(2) Der Hund jagt die Katze, die der Besitzer gerne streichelt.
   The dog chases the cat, whom the owner likes to pet.

In addition, it is possible to stack embeddings of relative clauses in so-called “center embedding”. See for instance Gibson (1998); Karlsson (2007); Lewis (1996) for examples and discussion.

⁹In German, it is also possible to for the relative pronoun to bear the role of indirect dative object (“Der Mann, dem die Frau den Kuchen bäckt, hat Geburtstag.”), genitive object (“Der Mann, dessen Frau ihn betrügt, ahnt nichts.”), as well as prepositional object (“Der Mann, von dem wir keine Entschuldigung mehr erwartet haben, kam gestern vorbei.”); prepositions in German can govern dative, accusative and genitive case.)—again, this list is not exhaustive. For a typological universal implication on the relativisation options available across and within languages, cf. the Accessibility Hierarchy by Keenan and Comrie (1977).
2. *Sentence structure complexity*

Analyses in head-driven phrase structure grammar (HPSG; cf. e.g. S. Müller, 2008) capture the two linkings with a mechanism similar to that used in accounting for unbounded dependencies, that is sharing of structure through so-called SLASH features which ‘percolate’ through a sentence structure (i.e. are transmitted via inheritance links) and mark the relative clause as ‘gapped’, i.e. containing an ‘extraction site’ for an unbounded dependency, calling for certain features to be selected. The selectional restrictions imposed by the SLASH feature are met by the formal features of the relative phrase containing the relative pronoun. In addition, a REL argument on the relative pronoun enables agreement features to percolate up to the mother node of the clause, resulting in a referential index for the entire clause, which again can be identified with the referential index of the noun it modifies in order to establish the link between ‘head’ noun and relative pronoun.

In transformation-based accounts, the relation between the modified head noun and the assumed relativisation site within the relative clause has been expressed with various formal tools, see Bianchi (2002a, 2002b), De Vries (2002) or Alexiadou, Law, Meinunger and Wilder (2000) for coverage of the different approaches. Bianchi (2002a, 2002b) for instance mentions matching plus deletion, where the relative clause is base-generated with a coreferential or identical NP, which is replaced by a relative pronoun. The pronoun is subsequently fronted and optionally deleted (Chomsky, 1965). Movement analyses in the wake of Chomsky’s conception of movement as copy-and-delete are akin to matching, only do they assume the relativised ‘head’ to originate within the relative clause. Such alternatives are also called ‘raising’ analyses, which assume that the matrix ‘head’ is inserted at the deep structure’s relativisation site and subsequently moved up (Bianchi, 2002a, 2002b). A still dominant approach of the latter type was proposed by Kayne (1994), as part of his theoretical treatment of antisymmetry phenomena (see however Borsley, 2001, for criticism of Kayne’s analysis).

With respect to a possible complexity difference between the structures used as stimuli in the present study, nothing much hinges on the technical details of how to formalise the relation between the relativised head noun in the matrix clause, the relative pronoun, and the ‘extraction’ or gap site within the relative clause. Similarly embedding as such might add to the structural complexity of sentences. I will not further cover center-embedding or ‘nesting’ phenomena in this work, as it proves quite difficult
2.2. Some structural problems

to disentangle the effects of syntactic embedding from the effects of additional information on the propositional level.¹⁰

Of more immediate importance to the present study is to account for the difference between subject and object relative clauses, and how this difference might be reflected in structural complexity differences. I will further restrict myself to subject-extracted relative clause structures, where for instance the difference between centre-embedding and right-branching or extraposition is not an issue:

(8) Hier bellt der Hund, der die Katze verfolgt.
Here barks the\textsubscript{nom} dog, \textit{relprn:nom;masc} the\textsubscript{f} cat chases.
Here the dog is barking, who is chasing the cat.

(9) Hier bellt der Hund, den die Katze verfolgt.
Here barks the\textsubscript{nom} dog, \textit{relprn:acc;masc} the\textsubscript{f} cat chases.
Here the dog is barking, whom the cat is chasing.

This type of sentences allows for an isolated structural manipulation. In comparison to for instance English, German SR and OR as in the examples 8 and 9 above do not differ in word order, but only in the case marking on the relative pronoun and the respective other determiner phrase, thus only in terms of a morphosyntactic contrast.

Structural accounts of the relative difficulty or complexity of SR and OR sentences have considered several factors to explain processing differences. Gorrell (2000) for instance suggests that the two sentence types differ in

¹⁰An obvious feature of relative clauses is that they are an instance of embedded (or dependent) clauses. The relative difficulty of structures with embedded clauses versus structures without has been in the focus of much linguistic and psycholinguistic research for decades, and it is widely assumed that embedding is a hallmark of greater structural complexity.

On the syntactic level, formal relationships between matrix and embedded clause need to be represented, where different types of embedding (e. g. complement, adjunct, or relative clauses) have been argued to differ in terms of processing difficulty, depending on the kind of relation that has to be processed (cf. Schütze & Gibson, 1999; Speer & Clifton, 1998; but also see Przepiórkowski, 1999 for criticism), and depending on the depth of embedding (Delage & Tuller, 2010). Hence, embedding of structure as such might accompanied by a higher amount of structure on the syntactic level. However, an increase in the amount of structure might go hand in glove with more propositional content, which needs to be processed. I will return to this problem in Chapter 3.
terms of the amount of structure or nodes necessary. In addition, the structural description of an object-topicalised sentence includes a dependency between the position where the bearer of the object role is expected and the position where it actually occurs. In the case of relative clauses, the distance between the relative pronoun and the site at which it can be interpreted (the verb or the pronoun’s trace position) has been argued to be longer in OR clauses than in SR clauses. Gibson (1998) for instance takes this as a key factor in his account on centre embedding, operationalising distance by counting discourse referents (nouns and verbs). What is more, in object relative clauses the dependency between relative pronoun and its ‘origin’ crosses a potential role filler. This opens up the possibility for an explanation in terms of an “intervener” configuration, which has been suggested for instance by Friedmann, Belletti and Rizzi (2009)—see Chapter 3, Section 3.4.2 for details on this account.

2.2.3. Agreement

Agreement is a special case of formal marking of dependency or concordance between two elements. In natural languages, agreement marks grammatical relations between elements of a sentence. In the case of agreement between verb and subject, the signalled relation is subjecthood, and since typically the subject bears the external argument role of a verb, the grammatical relation stands for a particular thematic or propositional configuration. This is realised as follows: an agreement target (e.g. the verb) bears identical values for one or more grammatical features as the so-called agreement controller (e.g. the subject of a sentence; Moravcsik, 2006), and this featural identity is signalled on a morphological level, for instance by the verb being realised in a particular form that represents a particular setting of feature values (e.g. 3rd person singular).

In many European languages including German, agreement can be found for instance for the relation between a subject NP and the verb. In German, lexical and auxiliary verbs share morphological marking for number with the subject noun phrase in person and number, to the effect that the verb bears an inflection ending which signals 1st, 2nd or 3rd person, and singular or plural number respectively. Nouns typically show inflection for number, with differences in terms of the morphological realisation between several noun classes.
The generalisation that elements in a sentence signal agreement by sharing features has led to formal linguistic descriptions that require a consistent unification of agreement features between for instance a subject noun phrase (NP) and a verb. This can be technically achieved by sharing of indices (cf. e.g. in HPSG: S. Müller, 2008; Pollard & Sag, 1994; Sag et al., 2003) or by unifying feature structures (Culicover & Jackendoff, 2005, cf. also Franck, Vigliocco & Nicol, 2002; Vigliocco, Butterworth & Garrett, 1996). Mainstream generative accounts differ slightly in their formal ways to capture agreement: The GB approach (Chomsky, 1981) formalised agreement in general terms based on the requirement for specifiers of a phrase in the X-bar scheme to agree featurally with its head (‘spec-head agreement’), and the assumption that the verb is incorporated by an abstract subject agreement head (see for instance Haegeman, 1995; Poole, 2002; also see Chomsky, 1995 or Radford, 1997 for a discussion of functional heads for agreement features like Agr[eement] or T[ense]). More recently, accounts have subsumed spec-head agreement under a generalised copying or checking relation ‘Agree’, cf. Grewendorf (2002) and references therein; also see Radford (2009). Under ‘Agree’, features of the highest node of a derivation (the ‘probe’, e.g. the subject noun), and of a suitable ‘goal’, for instance a verbal head are set reciprocally (Grewendorf, 2002; Radford, 2009).¹¹

Agreement appears as a rather widespread grammatical feature of many languages (including German), which at first sight does not add much to complexity on the structural level, especially when assuming a formal account which allows for an unbounded grammatical dependency like ‘Agree’. However, when agreement goes awry, the intricacies of establishing the correct formal link between two elements in a sentence become apparent. Previous studies have researched the role of different, mostly linguistic factors on the production of agreement attraction errors, as for instance in the following example 10:

¹¹Note that this newer conception based on the copying relation ‘Agree’ is conceptually very close to a constraint on agreement feature identity which licenses structure in HPSG or Simpler Syntax. An alternative construal of ‘Agree’ is that the verb is assumed to be pre-specified with agreement features in the lexicon and only ‘checks’ these features against those of the agreement head through movement, see for instance Radford (1997). Franck, Lassi, Frauenfelder and Rizzi (2006) propose a combination of feature copying under ‘Agree’ and subsequent (re-)checking of features in a specifier-head relation between the subject and the verb.
2. Sentence structure complexity

(10) (The inscription)\textsubscript{controller;sg} on (the ancient pillars)\textsubscript{attractor;pl} (*\textsubscript{were})\textsubscript{target;pl} weathered.

Already some of the earliest descriptions of attraction errors in the linguistic literature noted an effect of linear distance between agreement controller and an interpolated noun or 'attractor' on the likelihood of attraction errors, and consequently dubbed the effect 'proximity concord' (Hale & Buck, 1903/1966, cited after Francis, 1986; Quirk, Greenbaum & Leech, 1989). For structures with a single prepositional phrase modifier used here and in many earlier studies, linear precedence between candidate noun phrases seems to play a major role in the explanation of attraction effects during on-line processing (Fayol, Largy & Lemaire, 1994; also see Haskell & MacDonald, 2005). However, the explanation in terms of a simple function of linear distance was contested by a number of studies that stressed the importance of the structural relation that holds between controller and attractor. Studies comparing different types and depths of embedding of attractors yielded evidence that the strength of the attraction effect also depends on the hierarchical distance between attractor and target (Franck et al., 2002; Hartsuiker, Antón-Méndez & van Zee, 2001; Vigliocco & Nicol, 1998), and on the type of interpolated structure.

Franck et al. (2006) and Franck, Soare, Frauenfelder and Rizzi (2010) propose a formal account according to which the ‘Agree’ relation between two elements is more complex if an intervening elements is present. This structural account of agreement attraction phenomena links the higher error rates found with some configurations to principles of syntactic locality, in particular to the concept of ‘relativised minimality’ (Rizzi, 1990, 2004). As discussed in Sections 2.2.1 and 2.2.2, this concept has been suggested before as a relevant factor in explaining greater difficulty with different structural phenomena such as passive sentences or with object relative clause constructions.

Alternative formal treatments have captured agreement error phenomena with recourse to the so-called head feature principle (Culicover, 2009; Sag et al., 2003), which states that a phrasal projection XP shares the features of its head.\footnote{This constraint is realised in a variety of ways in different grammar formalisms, e. g. as ‘percolation’, copying, or identification/unification of feature sets (cf. Vigliocco et al., 1996).} A violation of the principle will result in an ungrammat-
2.2. Some structural problems

Apart from configurational properties like distance and locality, morphological factors can be shown to influence the correct establishing of agreement. In particular, formal properties of inflectional morphemes and the form of articles have been shown to play a role (e.g. Hartsuiker, Schriefers, Bock & Kikstra, 2003), as well as the amount of syncretism in morphological paradigms (Franck et al., 2002). Taken together, these results indicate that the structure of the morphological paradigm strongly bears on the likelihood of agreement errors to occur. As an example, Eberhard (1997) and others have reported “plural markedness” effects, characteristic for the elicitation of agreement errors in English: When the actual agreement controller is marked singular and a potential attractor with plural marking is present (see example 11), attraction errors are significantly more likely than in the opposite case (see example 12).

(11) The inscription on the ancient pillars *is/*are weathered.
(12) The inscriptions on the ancient pillar *is/are weathered.

This phenomenon can be theoretically described with a feature percolation approach, see for instance Eberhard (1997). The author claims that number is marked by a privative feature for plurality that can percolate to other nodes in the sentence structure. Singular nouns are assumed to lack this feature, which would explain the virtual absence of attraction errors when the local noun is marked for singular. This is in line with linguistic theory, as different authors have attributed plural number a ‘marked’ status over singular for phonetic, morphological, and conceptual or semantic reasons, see for instance Greenberg (1966/2005), Givón (1991), or Wurzel (1998)—see however footnote 1 on page 11 on the criticism by Haspelmath (2006) that the term ‘markedness’ might lack substance in many cases.

It should be borne in mind that languages differ in the way they code number, and morphological complexity does not necessarily reflect greater conceptual complexity or markedness. In particular, there may be more to the plural markedness effects than for instance the frequency of an individual form alone, as (language-specific) properties of the respective verbal paradigms and properties of lexical access come into play here (see for instance Kostić, 1991, also see Baayen, Levelt, Schreuder & Ernestus, 2007; Milin, Filipović Djurdjević & Moscoso Del Prado Martín, 2009; Van Ewijk
2. Sentence structure complexity

& Avrutin, 2010). In fact, Studies by Franck et al. (2002) and Franck et al. (2004) have pointed to cross-linguistic differences, since they report what appears to be a singular-markedness effect for French. The authors link these effects to the actual form of the inflectional paradigm, specifically to the amount of syncretism.

2.3. Linguistic complexity

The description of different structural phenomena considered as ‘complex’ from a syntactic point of view showed that very diverse factors have been suggested to be responsible for what is usually called ‘complexity’. In the following, I will discuss the concept of complexity from a more general point of view in order to frame linguistic definition attempts.

2.3.1. What is complexity?

As a general concept, the term ‘complexity’ is widely used in everyday situations as well as in scientific work. Intuitively, many uses refer to situations or entities which cannot be grasped ‘at a single glance’, or for which some considerable effort in understanding or explaining has to be taken. In his book on the mathematics of complex systems, Wolfram (2002) points to the fact that generally, concepts of complexity are tied to the goal of giving a short or simple description of something (or at least of some thing’s “features in which we happen to be interested”, Wolfram, 2002: p. 557). If this goal is not met, or seems unattainable to the observer, the object to be described is considered ‘complex’ by the person giving a description. Wolfram’s point is that some properties of an object invoke the impression of complexity to the observer, because these properties require an observer to use additional information for a description of the object (also see Härtl, 2008, for a similar view).

Attempts to formalise a generalised notion of complexity have proven difficult. Even in ‘exact’ natural sciences, in mathematics, or in philosophy it still appears quite poorly understood what kind of property ‘complexity’ might be (Wolfram, 2002). Formal conceptions of complexity have been outlined in other areas as well, including cybernetics or theoretical computer science (Jantzen, 1996; Kliegl & Fanselow, 1996). But in many areas even formalised accounts of complexity are based on the kind of general,
more or less intuitive notions of complexity that Wolfram describes, for instance as some sort of function of the amount of effort it takes to describe something. For instance, a machine can be considered relatively more complex than another one, the more explanation/description it takes for it to be operated properly.

It might not come as a surprise that Wolfram’s insights also hold for most if not all formal accounts of (linguistic) complexity. With respect to language, informal or intuitive definition attempts of sentence complexity usually refer to *representational* aspects of a sentence structure: if a sentence contains many words, shows unusual (or ‘marked’) word order, or conveys much information, it is considered ‘complex’. Before I will discuss formal notions of complexity in linguistics, I will first take another step back and consider what ‘formal’ means with respect to linguistic theory.

### 2.3.2. Formalising language

Linguists have been using formal descriptions of language as a tool for analysing the structure of sentences for decades, if not centuries (Givón, 1991). The term *formal* in this context is often used to simply mean ‘explicit’, that is with unambiguous terminology applied to phenomena (Chomsky, 2004). The sense of ‘formal’ implied here is basically related to the observable form of a particular element under scrutiny, for instance a word, and to the consciously performed abstraction step of forming *categories* (e. g. for the syntactic category of words). This categorisation in turn might be tightly linked to observations of formal similarities on other levels, for instance phonology or morphology and features on these levels again allow categorisation of observations. In addition, distributional properties of elements with respect to their context are taken into account to support the assumption of categories of elements (e. g. ‘natural classes’ in phonology; Hall, 2000). In this sense linguists have been formalising (categorising) language structure descriptions for quite some time, as soon as categories were identified, all as a means to bring “order to a given [empirical] domain” (Harman, 1980: 21).

The approach pioneered by Noam Chomsky in the late 1950s and early 1960s differed in that it combined linguistic theory for natural languages with mathematical approaches to formalise the treatment of string manipulation (*formal language theory*; cf. for instance Bertsch, 1989; Klenk, 1989;
2. Sentence structure complexity

Levelt, 1974/2008). Applications of the term ‘formal’ to linguistic theory in the aftermath of Chomsky’s innovation are thus closely linked to notion of a formal system as in maths or logic, which defines a finite alphabet and a set of mechanically applicable formation rules. This combination allows the contradiction-free generation of ‘well-formed’ sequences of signs (sentences), based on purely ‘formal’ criteria, i.e. irrespective of the (macro-level) meaning of a string. Such a combination of alphabet and formation rules is referred to as a generative grammar.¹³ A concise definition is given for instance by Uszkoreit (1996): “A generative grammar of a language can, in the broader sense, be considered a finite definition of the infinite amount of well-formed sentences for that language.” (p. 236).

For artificial languages like programming languages, a grammar is the collection of deliberately pre-specified formation rules that allow the generation of all well-formed sentences of a language, and only those. A grammar theory for natural languages differs from formal language theory as applied to artificial languages, however, in that the former is actually a profoundly empirical enterprise. Researchers cannot pre-conceive of the constraints that hold for natural language, instead they have to rely on theoretical reconstruction of observations from the domain of phenomena. Or as Levelt (1974/2008) put it: “observable phenomena correspond with that which is called LANGUAGE in the theory. The theory is an abstract description of the kind of system a natural language is.” (p. II:4).

Over the course of time, different formalisms for natural language grammar theory have been developed, which use different technological means to describe sentence structure. Early incarnations of the Chomskyan grammar theory used rewrite rules (based on a so-called ‘Semi-Thue’ system; cf. Pause, 1996), later generalised in X-bar theory’s assumptions about the form of a tree graph (e.g. binary branching, no crossing of edges, head is a sister to complement, etc.). These rewrite rules were extended

¹³In linguistic discourse, the term ‘generative’ has commonly taken on a more restricted meaning, usually referring to a specific flavour or incarnation of syntactic theory in the more immediate tradition of Chomsky’s work, cf. S. Müller (2010) on the terminology. In order to make a distinction between reference to the generative capacity of a grammar in general, and reference to more specific grammar theories, Jackendoff (2002) for instance uses the term ‘mainstream generative grammar’ (or ‘MGG’ for short) to denote the latter. For the sake of clarity, I will adhere to this convention as well and use the term ‘generative’ in its broader sense.
2.3. Linguistic complexity

with ‘transformational’ operations intended to extend the generative power of the formal language model and thus capture the semantic relationship between different surface structures (e.g. active and passive). The transformational operations have been successively generalised in mainstream generative frameworks like Government and Binding (GB) or the Minimalist Program (MP), and in current versions of the Minimalist Program they are subsumed under the general operation MERGE. In addition to the basic properties of the notational form, MGG theories contain grammatical principles (for instance about the locality of dependencies within a sentence), which constrain the transformational operations possible. Typically, the representational format of theories in the transformational tradition is that of a step-wise ‘derivation’. This format is also called ‘proof-theoretic’, or ‘generative-enumerating’, since the well-formedness of a sentence is derived in a form similar to a mathematical proof (Kiss, 2007; S. Müller, 2010).

Alternative accounts in the broader generative realm have also originated from phrase-structure grammars using rewrite rules, but to a large extent they avoided the introduction of transformations on the syntactic level (see S. Müller, 2010: 383), including variants of GB theory which do not use transformations (Koster, 1978, 1987, 2010; also consider the notion of ‘filters’, Chomsky & Lasnik, 1977), as well as theories like Lexical-Functional Grammar (LFG; Ronald M. Kaplan & Joan Bresnan, 1982), Generalised or Head-Driven Phrase Structure Grammar (GPSG, HPSG; e.g. Pollard & Sag, 1994; Sag et al., 2003), Tree-Adjoining Grammars (TAG; e.g. Joshi, Levy & Takahashi, 1975; Kroch & Joshi, 1985) or variants of syntactic theory formalised in Optimality Theoretic terms (OT; G. Müller, 2012; Prince & Smolensky, 1993/2002). Many newer variants of theories like LFG, GPSG/HPSG or TAG have been reformulated with the help of set theory, see S. Müller (2010): 59, and pp. 325–331. In these cases the grammar theory also postulates a formal apparatus for representation, and in addition it contains formalised constraints which hold for the set of all well-formed sentences (and only those). Crucially, constraints hold at the same time and exclude or ‘forbid’ structures which are not grammatical, rather than allowing only the grammatical structures to be derived.

Because many if not most aspects of syntactic analyses can be translated between theories, different formalisms are sometimes considered to be ‘notational variants’ of each other (cf. Chomsky, 2004; as a case in point consider for instance the formalisation of a Construction Grammar with
2. Sentence structure complexity

the HPSG formalism by Sag et al., 2003). An example of such aspects is the basic notion of ‘merger’, as used in MGG theories. Its analogon in other theories like HPSG or Simpler Syntax is ‘unification’, which has different mathematical properties (Jackendoff, 2007: 363), but can be seen as formally expressing a largely equivalent idea of combining more basic elements into larger structural contexts (Culicover & Jackendoff, 2005; Hagoort, 2005). Similarly, the concept of ‘movement’ as a formal representation of intra-sentential dependencies in transformational grammar formalisms has a non-transformational equivalent in so-called ‘slash’ or ‘gap’ features used in HPSG, which are inherited between levels of structure (Culicover & Jackendoff, 2005; S. Müller, 2008, 2010). Finally, functional topological patterns, i. e. the relationship between information structure or semantic/discourse functions and word order topology, have been captured with the help of so-called topological fields (Höhle, 1983/2003), as templates pairing form and function (or meaning) in variants of construction grammar (Goldberg, 1995, 2003), or in terms of functional projection hierarchies in so-called ‘cartographic’ approaches (Rizzi, 2004); see S. Müller (2010), p. 91.

2.3.3. Linguistic metrics

Formal descriptions of sentence structure in one of the different frameworks have been considered a useful starting point for the development of tools to evaluate linguistic behaviour (Lewis, 1996). Various kinds of observations have been made about differences between types or categories of syntactic structure. Observational studies about language acquisition and language breakdown, as well as corpus studies have indicated that certain sentence structures are used with considerably lower frequency of occurrence than others, or hardly occur at all. Research on typical language acquisition for instance has aimed to establish developmental pathways for comprehension and production ability in children, from ‘simpler’ to more ‘complex’ sentence structure (cf. e. g. Bloom, Lahey, Hood, Lifter & Fies, 1980; Brown, 1973; Rosenberg & Abbeduto, 1987). These pathways can be contrasted with acquisition patterns found in children with language impairments. In this vein, research on hearing impairment and

2.3. Linguistic complexity

language pathologies has yielded evidence that certain types of syntactic structure are more error-prone in production, in addition to being acquired later (Brannon & Murry, 1966; Delage, Monjauze, Hamann & Tuller, 2008; Delage & Tuller, 2010; Friedmann et al., 2009; Friedmann & Szterman, 2006; Hamann, Tuller, Monjauze, Delage & Henry, 2007; Svirsky, Robbins, Kirk, Pisoni & Miyamoto, 2000). In order to frame these observations, linguistic metrics have been developed which are used to assess behaviour based on formalised descriptions of sentence structure.

Starting from the intuitive notion of ‘complexity’ outlined in Section 2.3.1, a simple construal of linguistic complexity might be based on the amount of elements in a given domain of analysis, for instance word length in phonemes or morphemes, sentence length in words or text length measured in sentences, or in other combinations of elements with a given domain of inquiry. On the level of sentences, this notion is the basic rationale behind the **mean length of utterance** (MLU) measure that is used to approximate grammatical complexity in children (Brown, 1973)\(^{15}\), or a **mean clauses per utterance** (MCU) measure used for instance by Kemper, Kynette, Rash, O’Brien and Sprott (1989) to assess adult performance.

Some researchers have argued that the MLU might be of limited reliability to approximate speakers’ syntactic abilities beyond early stages of language acquisition (according to Rondal, Ghiotto, Bredart & Bachelet, 1987, for instance not beyond an MLU of 2.50), or have questioned whether the measure is valid for tapping specific morphosyntactic aspects of development at all, cf. Dethorne, Johnson and Loeb (2005). What is more, linguists have been describing intricate differences between sentence structure types which apparently affect how well a sentence can be processed or which influence how difficult a structure is considered in grammaticality or acceptability ratings. For such differences, measures like word count or MLU form rather coarse tools, see however Blake, Quartaro and Onorati (1993) or Szmrecsányi (2004), for different positions. This is apparent especially under *ceteris paribus* conditions where sentences do not differ in length; cf. for instance examples 6 and 7 from Section 2.2.2, repeated here for conveni-

\(^{15}\)MLU as originally conceived of by Brown (1973) counts the number of morphemes per utterance. In practice, the relatively time-consuming calculation based on morphemes is frequently replaced by a measure based on words per utterance (MLU\(_w\)). According to MacWhinney (2000), the two variants of the MLU measure show a reasonable correlation.
2. Sentence structure complexity

In those cases the complexity of a sentence is not simply a function of the amount of words it contains, but other factors have to be taken into account as well—most prominently among them the syntactic structure of a sentence.

Some examples of more fine-grained metrics as procedures for the analysis of sentence complexity of spoken language in special populations include the metric for assessment of agrammatic speech by Saffran and colleagues (Rochon, Saffran, Berndt & Schwartz, 2000; Saffran, Berndt & Schwartz, 1989), the D-level score proposed by Rosenberg and Abbeduto (1987) to account for acquisition pathways of children, or Scarborough’s Index of Productive Syntax (IPSyn; Scarborough, 1990). These approaches count occurrences of structure types considered ‘complex’ based on linguistic theory or intuition, usually without much further commitment to a specific (formal) linguistic theory (Hockl, 1996). An early suggestion to measure sentence complexity based on more explicitly formal assumptions about a language model was Yngve’s algorithm to measure branching depth of tree graph representations (Yngve, 1960). In a similar fashion, Johnson (1966) counted nodes and parsing steps through a tree graph as a measure for sentence complexity. In both cases (see Roark, Mitchell & Hollingshead, 2007, and Szmrecsányi, 2004, for more examples), general formal properties of the representation of a sentence in a tree graph structure are used as measures for complexity.

A slightly different approach which has been suggested by Lin (1996) and Gibson (1998, 2000) measures the length or distance of dependencies within a sentence. Their metrics are based on a formal description, but also takes aspects of meaning into account: Lin counts the total sum of the distance in words between heads and their respective modifiers, while Gibson measures the distance between an element and its attachment site.
2.3. Linguistic complexity

in a sentence structure, based on the amount of intervening elements that refer to new discourse referents, i.e. nouns and verbs.

In psycholinguistic research that is grounded in mainstream generative syntax, metrics more closely modelled on specific concepts from linguistic theory have been proposed. Examples are the Derivational Theory of Complexity (DTC; cf. for instance Berwick & Weinberg, 1983; Fodor & Garrett, 1967; Harley, 2001; Marantz, 2005; Phillips, 1996), or a more recent variant of the DTC, Jakubowicz’s ‘derivational complexity hypothesis’ (DCH; Jakubowicz, 1989, 2011). Both approaches provide complexity metrics that are closely tied to specific variants of a transformational generative syntactic theory (the ‘REST’ or the ‘Aspects’ model in the case of the DTC, and GB or the MP in the case of the DCH). In essence, both metrics define the complexity of sentence structure as a function of the amount of derivational steps necessary for the structural description of a sentence in the syntactic theory’s formalism. More specifically, in the case of the DCH the factor suggested to be crucial for increasing complexity is movement (‘internal merge’ in current MP terminology): a structure which involves more movement operations than another structure will be considered more complex.

Finally, an approach that combines aspects of the other methods has been pursued by authors like Delage et al. (2008); Delage and Tuller (2010); Hamann et al. (2007). Besides the theory-internal concept of ‘movement’, the authors’ metric takes other factors into account, for instance depth of embedding, or the number of violations of grammaticality constraints such as relativised minimality, or by measuring distances of intra-sentential dependencies, similar to Gibson or Lin.

2.3.4. Grammar theory and the goals of investigation

As I have tried to illustrate in the previous section, the range of linguistically-based metrics and their notions of sentence complexity is rather broad. A critical assumption that is implicit in a broad range of psycholinguistic work on sentence processing is that a structure which is more ‘complex’ at some level of formal linguistic description will also be cognitively more ‘costly’ to process (see e.g. Fanselow, Kliegl & Schlesewsky, 1999; I will return to this issue in Chapter 3 for a discussion of cognitive ‘cost’ could refer to).
2. Sentence structure complexity

However, despite their various degrees of success for diagnostic or therapeutic uses, so far none of the proposals can be considered unanimously accepted, or a ‘gold standard’ to capture complexity; see for instance Lewis (1996) for a review. What is more, especially metrics like the DTC, which closely operate with theoretical concepts from formal syntax theory, have been criticised by some researchers in psycholinguistics (F. Ferreira, 2005; Garrett, 1980; Harley, 2001; Levelt, 1974/2008; Slobin, 1966; also see Jackendoff, 2002, 2007 for discussion). In contrast to that, some authors consider the DTC and variants still state of the art metrics (Chesi, 2004; Franck et al., 2006, 2010; Jakubowicz, 2011; Marantz, 2005; Phillips, 1996).

As shown earlier, virtually all formal definitions of ‘complexity’ can be boiled down to a definition of the effort necessary to represent an observation formally. How the explicit, conscious process of creating a description for syntactic structure relates to implicit, unconscious processing remains unclear, however. In addition, we have to deal with the epistemological problem that the descriptive goals of linguistic metrics and the goals of the grammar theories from which they have been taken are not the same. While the latter usually try to model a distinction between grammatical and ungrammatical sentences (a binary distinction), the former have to specify criteria for modelling the gradual distinction between sentences which are easy or difficult to process.¹⁶ And while a metric is aiming to model processing behaviour, the empirical domain of grammar theories has been comprised to a large extent of the end-product of a production or perception process, by means of collecting data about native speakers’ intuitions about the ‘grammaticality’ of sentences (or actually their ‘acceptability’, cf. Bard, Robertson & Sorace, 1996; M. Meyer, 2006).¹⁷ In essence, grammar theory for natural language reconstructs intuitions and other evidence in the form of a language model (a ‘fragment’) and explicates rules or constraints that allow the prediction of the well-formedness of sentences added to the model (cf. Levelt, 1974/2008; M. Meyer, 2006).¹⁸

¹⁶It is beyond the scope of this work to enter the discussion about gradient levels of grammaticality or acceptability, but see for instance Fanselow (2006) for details.
¹⁷Levelt (1974/2008) describes the situation of linguistic theory in the early 1970’s, the time of writing his book, as “largely intuition-based” (p. P:9). The situation has changed in recent years, with evidence from corpus-based or experimental studies expanding the empirical domain of grammar theory.
¹⁸In addition to the reconstruction of empirical language data, structural properties are
A grammar theory for natural language then serves to predict the grammaticality of a sentence structure, presupposing that every human (native) speaker of a given language somehow observes or 'implements' the constraints explicated in the grammar theory. However, judging from the way it is usually built, a grammar theory is a theoretical, consciously performed abstraction of certain, select properties of (implicit, i.e. automatic and largely unconscious) linguistic behaviour or processing (Phillips & Lau, 2004; Vogel, 2009). This in turn means that any notion of sentence complexity which has been specified in purely linguistic terms will never be independent of implicit processing facts. In order to provide a formal apparatus that allows for a quantification of complexity/difficulty, a metric explicates these assumptions by either adding concepts external to linguistic theory (e.g. the concept of “memory units” in Gibson’s metric), or by re-interpreting properties of the grammar formalism the metric is based on in terms of a processing model (as e.g. in the case of the DTC, which posits that the formal derivation steps carried out for a linguistic analysis are somehow retraced during on-line processing).

The formalisations underlying linguistic theories as well as linguistic metrics, however, are based on observations by speakers and influenced by intuitive notions about what is deemed ‘normal’ language use and which types of sentences are ‘difficult’. Put differently, grammar theory for natural language is formally reconstructed from linguists’ observations about processing, which will be influenced by processing factors (cf. for instance Haider, 2007).

In sum, this means that linguistically-based notions of complexity are in fact not simply pertaining to an inherent property of sentences, but to processing difficulty operating over a particular kind of data. The properties of the processing mechanism we have at our disposal make us experience a particular data structure as ‘difficult’. In order to establish a meaningful concept of ‘complexity’, we need to know how the brain/mind processes different kinds of sentence structure. With respect to linguistic metrics, I follow Hockl (1996) in his assumption that metrics form measuring tools for both a processing mechanism as well as properties of a special type of data structure, without sufficient theoretical means to differentiate the two con-

---

deduced based on more general conceptual desiderata, like for instance conceptual economy, universality, or nativism.
2. Sentence structure complexity

cepts. The processing mechanisms necessary to produce and comprehend language might be operating very selectively or domain-specific, as for instance Franck et al. (2010) or Belletti, Friedmann, Brunato and Rizzi (2012) have argued; also see (Phillips, 2012) on this issue. Nevertheless, as I will argue in more detail in the following chapter (see especially the discussion in Sections 3.3 and 3.5), the subject matters (or explananda) of grammar theory and processing models belong to different theoretical domains, and the relationship between both has to be made explicit.

2.4. Summary

With this chapter I had set out to identify structural factors or properties which make a particular linguistic data structure more ‘complex’. However, the review of different conceptions of complexity on the linguistic level lead to the (perhaps unsatisfactory) conclusion that structural complexity described by means of a theory-based linguistic metric might be empirically and theoretically inseparable from processing difficulty. Given the goal of the present work, to explain observable behaviour with respect to the processing of different types of sentence structure, it is of high utility to have formal tools at hand to explicitly describe the types of data structure that might pose difficult for the human brain to process (cf. Harley, 2001). However, no formally-based metric or criterion for complexity is empirically independent of processing assumptions and hence sufficient to explain a particular complexity-related phenomenon. Such a metric allows to hypothesise about the processing mechanism only insofar, as it formalises what kind of data the processor has to process, thus formally setting “boundary conditions” (Jackendoff, 2002: 27) for a theory of processing. In essence, structural notions and processing assumptions form two sides of the same coin, and linguistic and psycholinguistic theory should go hand in glove. An explanation for why embedding or distance of dependencies lead to processing difficulties will come from a processing model that allows somewhat ‘realistic’ inferences about the individual processing steps happening in real time during the generation or parsing of complex structures.

Based on these considerations I will assume an essentially instrumentalist, pragmatic viewpoint on generalisations from formal language theory
2.4. Summary

when describing examples of ‘complex’ sentence structure used in the empirical part of this thesis.¹⁹ Linguistic criteria definable on the level of syntactic structure should be useful for predicting the relative processing difficulty of sentences in cases where different theoretical reconstructions ‘converge’ in terms of the complexity of the formal description. Accordingly, I will take such convergent predictions as a starting point for assumptions about the processing of these structures (see Chapters 3 and 5). This is a procedure which has in fact been followed by many linguistic complexity metrics and processing models in the past (cf. for instance Chesi, 2004).

¹⁹Following for instance Bender (2002); M. Meyer (2006), and Vogel (2009) I assume that grammar theory is presently still largely descriptive enterprise which is concerned with adequate structural analyses of sentences based on “patterns within and across languages” (Bender, 2002: 433; also see Blutner & Strigin, 2011; Koster, 2010; Kusters, 2008 on this issue; but see for instance Chomsky, 2004; Hauser, Chomsky & Fitch, 2002, who set out to reach more advanced goals).
3. Sentence processing-for-production

3.1. Introduction

The goal of the work presented here is to systematically test the influence of an environmental factor like noise on the production of complex or rather ‘difficult’ sentence structure. While the methodological approach of this work is exploratory, some initial model for language processing-for-production is necessary to frame the observations. In Chapter 2 I have discussed some of the representational properties of sentence structure which could make a particular kind of data difficult to process. I also argued that many formal linguistic metrics crucially rely on more or less implicit psycholinguistic assumptions about processing in order to predict or explain the difficulty experienced with certain types of sentence structure. While importing constructs from neighbouring theoretical areas generally is a welcome device to reach an explanation, such constructs or assumptions need to be made explicit. Otherwise the model or theory in question runs the risk of making pre-theoretical or simplistic assumptions, as for instance Levelt warned already in 1974:

"psychological presuppositions are implicit in the theoretical interpretation of certain linguistic observations. . . . One might wish that every linguist would fully recognize the role of psychological assumptions in the formulation of his theories. . . . At present psychological theory is implicitly working its way into the formulation of linguistic theory, instead of explicitly being taken into account and thus held in control. There is no reason to suppose that the common sense psychology of linguists is in any way better than the common sense linguistics of psychologists. (Levelt, 1974/2008, p. II:7)"
3. Sentence processing-for-production

In their work on language processing for both production as well as perception, many researchers in psycholinguistics follow the theoretical and methodological tradition of cognitive psychology, in that they assume the existence of mental states (which are considered inaccessible to empirical observation; Harley, 2001). A basic tenet of the tradition of the information processing approach pursued here is the view that cognition at some level of observation or modelling involves the manipulation of symbols. By following this tradition I subscribe to the so-called Computational Theory of Mind; cf. for instance Horst (2011) for an overview. Given the theory’s stance that the brain in essence is a form of computer, some form of mechanism or algorithm for language processing needs to be postulated, in order to explain the linguistic behaviour observed in the empirical parts of this work. Thus perhaps the most prominent goal of modelling in the information processing tradition in psycholinguistics has been to specify a logical and temporal order of processing steps, or to sketch the flow of data through the processing system (Horst, 2011; Levelt, 1989; Pechmann, 1994; Pickering & Garrod, 2004). Such a model should achieve psychological plausibility or external validity by exposing meaningful ‘interfaces’ with related theoretical domains from cognitive psychology, for instance memory psychology.

Because of the experimental focus of this thesis, the theoretical chapters will deal with existing models only, and I am going to pay special attention to the processing necessary for language production.¹ This chapter aims at explicating at least some of the assumptions that a useful model of language processing should contain, in order to allow for the following chapters to investigate possible loci at which the effects of noise and the processes involved in speaking interact. The last section presents some processing features relevant to the empirical part, and tries to link structural generalisations to processing mechanisms, if possible.

¹However, mostly for methodological reasons the amount of data on the time-course and on-line characteristics of language processing-for-production is much scarcer than comparable evidence about comprehension. In what follows, I will therefore deal with evidence from both production and comprehension of syntactic structure.
3.2. Levelt’s language production architecture

Among the most prevalent psycholinguistic theories about language production are the levels-of-processing accounts proposed by Victoria Fromkin (Fromkin, 1971), and Merrill Garrett (Garrett, 1980). These earlier accounts were based largely on speech error data, and have been extensively complemented by experimental data in later work. Researchers like Kathryn Bock (Bock, 1982) and most notably Willem Levelt (Levelt, 1989) elaborated earlier accounts into a comprehensive architecture for the language production system, sketched in figure 3.1 (also see Bock & Levelt, 1994 and Pechmann, 1994). As a result of experimental research in recent decades, parts of the theory have been further formalised and implemented as a computational model in a network of nodes with activation spreading between connected nodes (WEAVER++, covering mainly the lexical retrieval stage, Levelt, Roelofs & Meyer, 1999). With respect to the levels of processing assumed here, there is still an intense debate going on about both position and characteristics of interfaces between levels, but as a general framework for researching language production, the model is widely accepted.

Levelt’s model posits three major stages in the production process: (a) conceptualisation, (b) formulation, and (c) articulation. During the conceptualisation stage, the communicative intention or conceptual message is established, by setting both text-/discourse-level as well as information structure level communication goals. Levelt calls these two planning phases during conceptualisation macro- and microplanning, respectively. The resulting conceptual message is fed into the formulation stage, which again is divided into two sub-stages, grammatical encoding and phonological encoding. During formulation, syntactic structure is created, and lexical representations for the planned utterance are retrieved and are inserted into slots of the syntactic structure. According to the model, processing on the grammatical encoding stage is further subdivided into functional and positional processing, the former creating hierarchical dependencies between elements of a sentence and assigning grammatical functions, and the latter subsequently creating linear order (Bock & Levelt, 1994; Garrett, 1980). The result of the formulation stage is a phonetic plan that is eventually translated into a motor plan and executed during the articulation stage.
3. Sentence processing-for-production

Figure 3.1.: Language production architecture, adapted from Bock and Levelt (1994) and Pechmann (1994).

3.2.1. Incrementality, seriality, and automaticity

The architecture of the Levelt model is broadly modular, and a serial flow of information is assumed, allowing for (strictly feed-forward) cascading of information between levels of representation (Pechmann, 1994). This way, parts of finished structure can be moved on to the following level for further processing, ensuring fluency. Hence a crucial property of the language production model sketched here is the assumption of incremental processing of utterances, i.e. that sentences are constructed in a piecemeal fashion from left to right (cf. F. Ferreira & Henderson, 1998; Kempen & Hoenkamp, 1987; also see Bock, Irwin & Davidson, 2004; Bock, Irwin, Davidson & Levelt, 2003; F. Ferreira & Swets, 2002, for discussions of the extent of incrementality in speaking).

Especially the amount of feedback between levels of processing has been
3.2. Levelt’s language production architecture

debated since the earliest days of psycholinguistic modelling of language production. While the model by Levelt (1989) and even more so the computational model by Levelt et al. (1999) by its very architecture allows for parallel processing and horizontal feedback only within one level, e.g. lexical retrieval, the flow of information between levels is assumed to be serial and top-down. However, the question of (vertical) seriality of information flow has been called into question for instance by connectionist, parallel processing accounts, with some implementations allowing for backwards information flow between levels of processing (Dell, 1986; Dell, Chang & Griffin, 1999). These accounts model processing without strictly separated representational levels in simulated networks of nodes with symbolic or sub-symbolic content, cf. Schade (1992). Vigliocco and Hartsuiker (2002) present a review of existing frameworks and propose a model that allows limited feedback and a moderate parallelism. According to the authors, the available evidence speaks neither in favour of minimal input nor in favour of unidirectionality of processing, and they propose as an alternative a competition model.²

Intuitively, language use seems like an easy task to (unimpaired) speakers, and is considered only partially accessible to consciousness. Under many circumstances, processing for both comprehension as well as production is considered to occur in a highly automatic fashion (Levelt, 1989). Because of this impression, language and especially processing of syntactic structure has been viewed as example cases of modular or informationally encapsulated processing stages, in the sense of Fodor (1983).

An obvious exception to the alleged automaticity in speaking is the conceptualisation phase in Levelt’s model, which involves the (conscious) selection of communicative goals and the assembly of a message. What is more, as the discussion by F. Ferreira and Engelhardt (2006) shows, there is mixed evidence about the question whether the processes involved in the assembly of syntactic structure might in fact be resource demanding,

²In this work I will mainly stay within the realm of ‘classic’ computational models that describe cognition as a system that manipulates symbols. The sole reason for this is simplicity of the modelling, as a connectionist implementation of the modelling assumptions made in this thesis is beyond the scope of this project. I do expressly not assume that the two modelling strategies are incompatible (cf. e.g. Garson, 2012), especially since aspects of connectionist modelling have been used in hybrid models such as WEAVER++ (Levelt et al., 1999).
and whether these demands vary with the complexity of the sentence that is about to be produced. Thus, the hypothesis of the apparent ease of use and automaticity of language production needs to be qualified: what does it mean to say that a cognitive process like language production appears to happen ‘automatically’, and to what extent is this actually the case? I will take up this question again in more detail in Chapter 4.

3.2.2. Monitoring

As stated in the introduction, speakers make errors, which have formed a valuable source of information about language processing. In fact, as F. Ferreira and Engelhardt (2006) argue, errors in production are more prevalent than usually thought. This fact prompted Levelt (1983, 1989) to include in his language production model a mechanism to check for errors on-line, the so-called monitoring system. This way Levelt’s model allows for limited feedback between processing stages, while still maintaining a broadly serial, feed-forward processing architecture.

The monitoring system as proposed by Levelt is assumed to ‘observe’ different aspects of language production, including “postural” (Postma, 2000), articulatory factors that influence phonation and fine phonetic detail, as well as conversational setting, (pragmatic) appropriateness, syntactic and morphological well-formedness and completeness. During the production process, the system compares different levels of structure down the line to the intended conceptual message. In case of mismatch, formulation and/or articulation can be interrupted and a repair can be attempted. According to Levelt (1983) and Levelt (1989), the monitoring component forms a means to limit the amount of errors made during production (Levelt, 1989: 460), and thus serves a “corrective” function (Postma, 2000).

In order to avoid a duplication of procedural (including grammatical) knowledge for monitoring components, Levelt proposes the perceptual loop theory, stating that monitoring makes use of the language comprehension system, with two possible access points at which representations of the production system can be monitored. The perceptual loop theory assumes two different channels through which (parts of) an utterance can be fed back into the comprehension system. Evidence from fast self-interruptions and speech-error-induction experiments yielded the notion of internal speech being monitored, the first available monitoring channel. An internal loop is
3.2. Levelt’s language production architecture

created in the model by allowing a direct connection between production and comprehension tapping the phonological plan or internal speech that is output from the formulation stage (Levelt et al., 1999). The external loop is the second channel included in the model. It draws on overt speech that is processed through the auditory perceptual system first and then fed into the comprehension system. A couple of other loops have been proposed since by other authors, and there has been an intense debate about the nature of the monitor: serial-modular, with access to many internal levels of processing, or production-system internal (see Postma, 2000 for a review of as many as 11 monitoring pathways within the language production system).

3.2.3. Embedded grammar

In order to capture and formally represent generalisations about sentence structure in the production process, a language production model needs to contain an implementation of a syntactic theory (Levelt, 1974/2008). In the case of the model by Levelt (1989), this is a variety of lexical-functional grammar (LFG). In particular, he chose a variant adapted to the needs of a production model that aims to characterise the time-course of building sentence structure in a piecemeal fashion: the incremental procedural grammar (IPG) by Kempen and Hoenkamp (1987); compare also the newer Performance Grammar theory proposed by Kempen and Harbusch (2002).

However, as for instance Pechmann (1994) stresses, it is difficult if not impossible to identify the one ‘true’ grammar theory to be integrated in a model of language production, given the multitude of formal accounts available (cf. Section 2.3.2 of Chapter 2). In the past, psycholinguistic accounts of processing phenomena have included various amounts of generalisations from formal linguistic theory. While some authors have sought to establish a close connection between theorising in linguistics and psycholinguistic modelling, others have remained agnostic with respect to the relationship between linguistic and psycholinguistic theory, and some researchers have explained sentence structure processing phenomena without recourse to syntactic theory (cf. for instance F. Ferreira, 2005, who bemoans

---

*See Oomen and Postma (2001) for a discussion of this proposal and its implications about the alleged modularity, in particular the assumption of informational encapsulation, of the production and perception system.*
an ‘alienation’ of psycholinguistics and linguistics since the demise of the derivational theory of complexity). Thus the challenge all psycholinguistic models of sentence processing for both production as well as comprehension are facing is to implement the constraints of grammar theory by modelling mechanisms for a real-time processing device with a certain degree of independence from theoretical developments, while still remaining theoretically compatible to linguistic accounts.

3.3. Linguistic and psycholinguistic modelling

A long history of psycholinguistic research has dealt with the question of how sentence structure considered ‘complex’ from a theoretical point of view is processed in real time during production and comprehension. For different linguistic factors believed to be indicative of higher complexity it has been investigated how they influence processing, these factors including among others: word order/canonicity (e.g. active/passive, topicalisation), amount and distance of dependencies (e.g. questions, topicalisation, relative clauses, concordance), (depth of) embedding, attachment and syntactic function ambiguity, interference between possible gap fillers or concord/agreement partners; cf. among others Bader (1994); Bader and Meng (1999); Fanselow et al. (1999); F. Ferreira (1991); Gibson (1998); Gordon, Hendrick and Johnson (2001); Konieczny (2000). The body of experimental work yielded some stable correlations between processing difficulties and theoretical constructs that aim to capture structural complexity (F. Ferreira, 1991; Gibson & Thomas, 1999). However, the relationship between theoretical constructs and processing evidence appears to be not as coherent as some studies have initially promised, see for instance Carrithers (1989); F. Ferreira and Engelhardt (2006) for critical discussion of the influence of frequency, or context conditions (also see Mak, Vonk & Schriefers, 2008; Weskott, 2002 for evidence to that end). Given this incomplete linkage of formal linguistic constructs and experimental evidence, researchers have been asking the highly interesting (although still unanswered) question how generalisations from structural accounts can be related in a meaningful way to real-time processing models for language production and comprehension (Garrett, 1980; Jackendoff, 2007; Levelt, 1974/2008; Phillips, 2012; Uszkoreit, 1996).
3.3. Linguistic and psycholinguistic modelling

3.3.1. Grammar theory and processing mechanisms

Since the inception of generative transformational grammar in the early 1960s it has been tempting for linguists and psycholinguists alike to interpret formal tools used by syntacticians to describe sentence structure as a mechanism for language processing (cf. e.g. Harley, 2001; Jackendoff, 2002; also cf. the discussion of linguistic metrics in Sections 2.3.3 and 2.3.4). As Berwick and Weinberg (1983) put it, under such an interpretation “the principles employed to describe the system of knowledge [emphasis added] that makes up the language faculty should also provide an adequate description of that system’s implementation in language use.” (p. 3). This position manifested for instance in the form of the so-called derivational theory of complexity, or DTC for short; other authors have referred to this assumption as forms of iso- or homomorphism between linguistic and psycholinguistic theory, cf. for instance Levelt (1974/2008) or Berwick and Weinberg (1983). More recent variants of the theory have been proposed for instance by Jakubowicz (2011), in the form of her ‘derivational complexity hypothesis’ (DCH), or by Franck et al. (2006), who consider the DTC a “null hypothesis”; also see Marantz (2005) for a defence of the DTC in reply to Jackendoff, 2002, and other critics of this approach.

Another recent example of an approach that postulates massive isomorphism between the parser and a grammar in language processing-for-comprehension is the work by Chesi (2004). His work is an attempt to construct a psychologically plausible model for syntactic computation in real time, based on a derivation-based grammar formalism (see Phillips, 1996, for an earlier approach in the same vein). Chesi embeds his model in considerations of mathematical complexity theory borrowed from theoretical computer science,⁴ which postulates the two concepts of time and memory re-

---

⁴The notion of computational complexity has been playing an important role as a concept in theoretical computer science for decades. The field has for a long time been dealing with issues of computability and complexity of problems and algorithmic solutions to problems. Computability theory deals with the question which problems can be computed at all with generalised machines, and classifies problems as ‘solvable’ (decidable) or ‘unsolvable’ (undecidable), see for instance Hromkovič (2011). According to Hromkovic, the issue of computability in fact touches the very philosophical foundations of computer science and mathematics.

The related field of complexity theory deals with the question how much resources are consumed for the computation of a particular decidable problem, given a particu-
source complexity of a problem being computed on a generalised machine model. The two central constructs of time- as well as resource-consumption from complexity theory have been considered useful concepts to deal with complexity issues in cognitive psychology, and Chesi (2004) argues that a derivation-based grammar formalism can provide the blueprint for a mechanism. As a result, he treats the formal process of coming up with a ‘derivation’ as an algorithm to estimate resource or time complexity of processing for judging, interpreting or generating a sentence.

As such, the assumption that theoretical concepts or constructs, (elements of a theory which are not accessible to direct observation, cf. e. g. Bortz & Döring, 2006) bear a more or less direct relation with an external reality forms a basic tenet of scientific realism (Boyd, 2010). Especially in cognitive psychology the notion of theoretical constructs has seen a great deal of methodological discussion (Bortz & Döring, 2006); and for a construct to be considered scientifically useful and in some sense or other ‘realistic’, methodological criteria have been established for the validity of constructs. The question then is how a realistic or isomorphic interpretation of a syntactic theory formalism as a processing mechanism fares with respect to different validity criteria. Some problems have been raised about this view, making DTC-like approaches look “misleading” (Jackendoff, 2002) or “unappealing” (F. Ferreira, 2005) to many psycholinguists ever since the demise of a machine model (some kind of generalised computer, typically a deterministic finite automaton like a Turing machine). The theory allows for a sub-classification of the solvable or decidable problems, depending on the amount of resources necessary for a solution given an ‘optimal’ algorithm (e. g. classifying a problem as solvable but requiring nearly infinite amounts of time). For this purpose, the notion of complexity has been explicitly formalised in terms of the time-complexity, postulated as the number of base operations necessary to solve a problem, and the spatial complexity, postulated as the amount of memory used for the calculation. In practice, a problem is classified in terms of its resource or time complexity (a) in the limit, (b) with respect to a generalised, abstract machine model that is assumed to be an adequate model for the architecture of the kind of computers we use today (Von-Neumann architecture; Hromkovič, 2011), and (c) with respect to the notion of an ‘optimal’ algorithm defined in the (programming) language the abstract machine model provides. The machine model describing this mechanism is selected according to the machines that (at least potentially) exist in the real world.

See however Kliegl and Fanselow (1996) for the stance that, in contrast to computer science, a similarly unified conception of complexity so far does not exist for the cognitive sciences, including linguistics.
of the DTC in the 1970s.

As for instance Bender (2002) argues, the interpretation of a grammar theory and its formal tools as a psychologically real mechanism that operates in real time has to be considered problematic, because it forms a reification or naturalisation of metaphors in the object of study. First, as mentioned before, currently a variety of different formalisms exist, but apart from the comparison of theoretic description with observable (processing) data, no principled criteria have been established to determine which one is ‘best’. Fanselow (2009) for instance bemoans a “certain arbitrariness in model building” within the generative mainstream. Given this situation, a whole array of different processing mechanisms are thinkable that can result in a complex or difficult grammatical structure as the end-product. Hence, postulating an isomorphic implementation of a theory’s formalism for the description of sentence structure as a mechanism for language processing means to ignore what Hilary Putnam called the problem of “multiple realizability” (Bickle, 2008). Jackendoff (2002) issues the warning that the investigation of actual brain circuitry and of the mind’s functional organisation might show that for processing language the mind “resorts to ‘cheap tricks’ rather than the mathematically most robust solution.” (p. 22).

What is more, data about the empirical adequacy of an approach like for instance the DTC or its newer variant DCH presents a mixed picture. While predictions about structural complexity derived from the theory coincide with observable behavioural differences in some cases, in other cases they do not (Belletti & Chesi, 2011; Carrithers, 1989). Jakubowicz’s DCH for instance postulates that all movement operations are cognitively costly. Without further qualification the hypothesis would predict that for instance in Germanic languages with V2⁸, movement of the verb from its sentence-final position to the position following the subject incurs costs.⁹ In effect, for languages like German, the metric makes the awkward prediction that processing of SVO should be more difficult than processing SOV structure. According to Fanselow et al. (1999) this is not likely to be the case (see

---

⁶Cf. M. Meyer (2006) for reasons why the theory-internal adequacy criteria proposed by Chomsky are of limited utility in this respect.
⁷“eine gewisse Beliebigkeit der Modellformulierung”, (p. 138)
⁸I.e., SOV word order in subordinate clauses and SVO word order in main clauses.
⁹See for instance Matushansky (2006) for a review and discussion about the status of head movement in the minimalist program.
3. Sentence processing-for-production

however Bastiaanse & Thompson, 2003, for some evidence to this end from speakers with Broca-type aphasia).

For these reasons the ‘mechanism’ suggested by a realistic reinterpretation of a particular grammar theory must remain purely metaphorical. In the light of these considerations, I assume that to this date the best view is the one already expressed at the end of Chapter 2, the view that a grammar theory is “epistemologically innocent”¹⁰ (Kiss, 2007) in terms of the processing device implementing it. This has been more or less explicitly acknowledged by grammar theories like some OT-implementations or variants of MP grammars which have been designed as a random generator, evaluating a potentially infinite amount of candidate structures with respect to ‘optimal’ empirical coverage of grammaticality/acceptability (Kusters, 2008; Prince & Smolensky, 1993/2002). As (Kiss, 2007) rightly points out, a random generator is a highly implausible model to be implemented as a mechanism in the brain. In fact, Chomsky’s own current position seems to reflect these reservations against the naturalisation or reification of theoretical concepts: in a 2004 publication he argues to view the derivational format as purely metaphorical (Chomsky, 2004).

However, as already concluded at the end of Chapter 2, this non-naturalistic view leaves us with two distinct theoretical domains of grammaticality and processing difficulty, and the relationship between the two appears to remain unclear. As the work by cognitive psychologist David C. Marr on human vision shows, this need not be a problem, because an ‘explanation’ of cognitive phenomena can only be attained by taking different perspectives into account anyway.

3.3.2. Marr’s Three-Tier-Hypothesis

As Marr and Poggio (1976) and Marr (1982/2010) have argued, different levels of description have to be distinguished when dealing with cognitive systems (also see Jackendoff, 2002, for more recent discussion based on Marr’s argumentation). With his so-called three-tier hypothesis, Marr raises a fundamental epistemological question about what constitutes an explanation in cognitive science—exemplified by his psychological theory of vision (Marr, 1982/2010). According to Marr “[f]or the subject of vision, there is

¹⁰“epistemologisch unschuldig”, (Kiss, 2007)
no single equation or view that explains everything. Each problem has to be addressed from several points of view” (p. 5). In essence, Marr’s stance is that a reduction of theoretical descriptions on one level of observation to a description on another will either be impossible or it will not necessarily provide an explanation, because the vocabulary used on the two levels is not compatible.

The merit of Marr and Poggio (1976) and Marr (1982/2010) is to have provided a framework for theoretically separating the levels of investigation in the vision system and identify mainly three levels of description, which taken together might allow for an explanation of cognitive functions:

The most abstract is the level of what the device does and why. . . In order that a process shall actually run, however, one has to realize it in some way and therefore choose a representation for the entities that the process manipulates. The second level of the analysis of a process, therefore, involves choosing two things: (1) representation for the input and for the output of the process and (2) an algorithm by which the transformation may actually be accomplished. . . If the first of our levels specifies what and why, this second level specifies how. . . This brings us to the third level, that of the device in which the process is to be realized physically. (Marr, 1982/2010: pp. 22)

As Marr (1982/2010) and, following Marr, Jackendoff (2002) have argued, the three-tier-view generalises nicely to other aspects of cognition, like language. And in fact, a separation of levels of investigation has been part of linguistic theory for quite some time.

3.3.3. Competence and performance

The three-tier hypothesis closely resembles the distinction between linguistic competence and performance as it was originally proposed by Chomsky.¹¹ However, the relationship between competence and performance

---

¹¹This relationship is explicitly acknowledged by Marr:

Chomsky’s (1965) theory of transformational grammar is a true computational theory in the sense defined earlier. It is concerned solely with specifying what the syntactic decomposition of an English sentence should be,
theories in sentence processing has been a highly problematic issue since
the inception of the two theoretical constructs: as (among many other au-
thors) for instance Levelt (1974/2008) and (Jackendoff, 2002) have argued,
the distinction between competence and performance has been put into
practice somewhat arbitrarily in the past. According to Jackendoff, the
distinction was originally conceived of by Chomsky as a matter of prag-
matically delineating the empirical domain of linguistics, but clear criteria
for separating the two domains have been lacking: “what Chomsky lumps
into performance actually constitutes a wide variety of phenomena. Some
fall into basic facts about memory limitations; some into different aspects
of the theory of sentence processing; and some now are subsumed under

It has been argued that some structural properties of language, which
have been captured with a theory-internal grammatical ‘principle’, might
fall naturally from (domain-)general properties of human cognition, more
specifically from properties of for instance human memory (Hawkins, 2004;
Lewis, 1996; Phillips & Lau, 2004). Apparent similarities between the pro-
cessing of linguistic and other kinds of structure (e. g. visual patterns)
might point to general principles of cognitive processing as it is implemen-
ted by the human brain. In this case the burden of explanation shifts from
the computational level to the algorithmic level. I wish to stress that this
does not preclude the existence of linguistic principles that fall from gen-
eral mathematical properties of computation: especially for the latter type
of generalisation, the only way to adequately capture them so far will be
with the help of a linguistic competence theory (again see Phillips & Lau,
2004 for a more detailed argument). It is to this day an empirical question
which aspects of linguistic behaviour are best captured with a competence
theory, and which are due to processing characteristics (Chesi, 2004; Levelt,
1974/2008). Such considerations have obviously influenced theory devel-
oped within the ‘mainstream’ Chomskyan tradition, as recent models of
the human language faculty show: grammar-specific processing or compu-
tation (the “faculty of language in the narrow sense”, Hauser et al., 200212)

¹²Note however that the authors are making an evolutionary argument. Also see the en-
are constrained by domain-general properties of cognition, like memory
capacity (see also Berwick, Friederici, Chomsky & Bolhuis, 2013).

The generally rather unclear nature of the relationship between levels has
been a problem not only for linguistic theory,¹³ but was also recognised by
With respect to the different levels of description proposed by Marr, a re-
ification of formal concepts in an algorithm forms a contamination of levels
which he calls a source of ‘confusion’ (Marr, 1982/2010: 25; also cf. Levelt,

Marr’s conceptual framework is highly useful to this day, since it allows
us to differentiate here: a syntactic theory which sees itself as a pure com-
petence theory in Chomsky’s original sense (and thus a computational/
mathematical theory in Marr’s sense) does not make processing assump-
tions in the same way a psycholinguistic theory does. Hence, a grammar
theory can tell us something about what the brain is able to compute; which
is a crucially different question than asking how the brain computes struc-
ture (Jackendoff, 2002; Marr, 1982/2010; Vogel, 2009). In order to avoid con-
tamination problems, a psycholinguistic concept of ‘computation’ needs to
be substantiated on an algorithmic level. This requires theoretical and em-
pirical links with neighbouring disciplines like memory psychology to be
established (cf. Lewis, 1996). I will return to this question in Chapter 4.

3.4. Processing of different phenomena

Several authors of linguistic metrics have argued that observable perfor-
mance differences can be explained with respect to processing resources and

壹³Consider Chomsky’s original statement on the relationship between competence and
performance:

Linguistic theory is concerned primarily with an ideal speaker-listener, . . .
who is . . . unaffected by such grammatically irrelevant conditions as
memory limitations, distractions, shifts of attention and interest, and errors
. . . in applying his knowledge of the language in actual performance. . . .
Only under the idealization set forth in the preceding paragraph is perform-
ance a direct reflection of competence. (Chomsky, 1965, p. 3–4, cited after
capacity limitations of the processor when handling ‘costly’ structure. With respect to the sentence structure types discussed in Section 2.2 of Chapter 2, I have argued that a complexity metric based on purely formal criteria is never free of implicit processing assumptions, and typically explanations of processing phenomena provided by metrics contain more or less explicit auxiliary assumptions. As argued in the previous section, processing ‘cost’ has to be substantiated by means external to linguistic theory, since a competence theory by definition abstracts away from many factors that pertain to processing.

As to what kind of ‘cost’ is actually incurred during on-line processing, accounts differ. Schlesewsky, Fanselow, Kliegl and Krems (2000) postulate memory load as storage cost (‘space’; but see Gorrell (2000) for discussion). The explanation by Friedmann et al. (2009) rests on syntactic working memory, without further specification of the concept or the ‘unit’ in which they would measure processing cost.\(^{14}\) Gibson (1998, 2000) postulates memory cost in terms of energy units that are used for both storage and integration.\(^{15}\) The parsing model by Gorrell (1995, 2000) remains largely uncommitted in this respect, but it seems possible that processing time could be implied as the unit. Finally, the working memory retrieval model by Lewis and Vasishth (2005) also postulates ‘time’ as the central unit of processing cost, based on insights from the ACT-R theory for cognitive processing (Anderson et al., 2004).

In the following, I will briefly review some processing accounts for the three different structure types researched in the empirical section of this thesis, and I will try to analyse the accounts with respect to the level of description (in Marr’s sense) they pertain to. An important aspect of this task is to identify plausible mechanistic/algorithmic concepts and check whether they can be substantiated in an independent fashion, or whether they are in fact reified metaphors. In addition, I will aim to explicate the concepts of processing ‘cost’ contained in the different accounts and provide some assumptions as to why a particular type of structure is difficult or ‘costly’ in processing.


\(^{15}\)However, definitions of memory cost in terms of an energy unit have been criticised for instance by Rummer, Mohr and Zimmer (1998), because such concepts fall short of explaining the processing mechanisms constituting (verbal) working memory.
3.4. Processing of different phenomena

3.4.1. Active and passive

A number of experimental studies has tried to establish measurable differences between the processing of active and passive sentences, and tested the claim that the structural differences between active and passive sentences would incur processing difficulties. Evidence from early studies on comprehension appeared to speak in favour of the hypothesis (cf. e.g. Forster & Olbrei, 1972; P. Gough, 1966; P. B. Gough, 1965; Mehler, 1963; Wang, 1970). However, a number of other studies have provided evidence that the argument is not as straightforward as it initially seemed. A number of non-syntactic factors seem to crucially influence the course of on-line processing, such as the lexical content (Matthews, 1968), semantic reversibility (Slobin, 1966), conceptual focus or ‘perspective’ (Olson & Filby, 1972; Tannenbaum & Williams, 1968), and conceptual accessibility (Bock & Warren, 1985). Carrithers (1989) criticises some of the earlier studies that focused on structural differences as not measuring data that pertain to syntactic processing itself, but rather to secondary tasks relying on semantic interpretation of passive sentences.

Especially the earlier studies by Matthews, Slobin and Olson and Filby play into the context of the early debates around competence and performance and the DTC. The accounts differ in terms of whether explicit reference is made to more general cognitive processing factors like working memory (see e.g. Matthews, 1968), but usually converge on the assumption that transformational complexity as modelled by grammar theory at the time cannot serve as the only factor explaining processing difficulties. Most of the authors cited above do not model a mechanism, which makes it difficult to assess the level according to Marr’s hypothesis. An exception might be Bock and Warren (1985), who place their results within the framework of psycholinguistic language production theory current at the time, and therefore can be said to model at the ‘algorithmic’ level.

More recent psycholinguistic research has featured active and passive structures in work on syntactic priming or structural persistence, i.e. the tendency to reuse structures that have been previously heard or produced (Bock, 1986). It has been observed regularly that passive structures tend to occur with lower frequency under baseline conditions than active sentences (Hartsuiker, Pickering & Veltkamp, 2004).¹⁶ Hartsuiker and Kolk

¹⁶‘Baseline’ here refers to cases where only active primes are presented, but an obvious
3. Sentence processing-for-production

(1998) for instance show strong priming effects for Broca-type aphasics on passive sentences, and interpret their results as evidence for a processing account of agrammatic disorders that can be temporarily overcome through priming.

Many accounts of syntactic priming assume lexical representations of syntactic information with fluctuating activation levels that can be influenced by exposure to the respective structure, see for instance Hartsuiker et al. (2004); Levelt et al. (1999); Pickering and Branigan (1998). Under this construal, the probability of a construction (or the baseline activation level, in terms of the model) serves as an explanation for speakers’ preferences for one over the other structure.¹⁷ A sensitivity to frequency appears to be a property of the processing mechanism implementing the selection of (partial) structure templates if alternative structure types are available to express a conceptual message.

As I have discussed previously in Chapter 2, one of the most striking structural aspects of the passive is the fact that arguments or role-bearers are not in the linear order they typically (or ‘canonically’) have, i.e. in the most frequent structures in a particular language. This factor has been taken as a hallmark of complexity/processing difficulty from both linguistic viewpoints as well as for psycholinguistic models. According to for instance Fanselow et al. (1999), what is particularly costly in ‘non-canonical’ sentences is the dislocation of arguments and the concomitant difficulty in assigning the right theta-role to the right argument. Drai and Grodzinsky (2006) argue in a similar fashion for why passives might be more difficult to process for patients with aphasia in English, but not in German: according to the authors, the important difference is movement of an argument across the verb (p. 163; see however Bastiaanse & van Zonneveld, 2006 for a criticism of Drai and Grodzinsky’s view).

Crucially, what the accounts by Drai and Grodzinsky (2006) and others boils down to is the fact that the order of argument-bearers in a passive sentence is different from what is assumed about the order of arguments

¹⁷Based on a different concept of cognitive ‘computation’ altogether, connectionist or parallel distributed processing accounts of structural priming have been proposed as alternative to symbolic activation spreading models (e.g. Chang, Dell, Bock & Griffin, 2000).
stored in a verb’s lexicon entry (e. g. Jackendoff, 2002; Kratzer, 1996). Many functional reasons for diverging from the ‘base’ order have been related to semantic factors, including for instance animacy (F. Ferreira, 1994; Grewe et al., 2006; Mak, Vonk & Schriefers, 2006), to discourse and information structure (Gordon & Chan, 1995; Mak, Vonk & Schriefers, 2002; Mak et al., 2008; Weskott, 2002),¹⁸ to more general cognitive factors like attentional focus (Tomlin, 1997) or conceptual salience (Bock & Warren, 1985; Prat-Sala & Branigan, 2000), or to specifics of lexicon organisation, like the availability of entries (V. S. Ferreira, 2003; Kolk, 1995; Levelt, 1989). Whatever the reason, however, a clear effect on the syntactic representation of the passive compared to active sentence structure is the necessity for additional structure.

As mentioned previously in Chapter 2, a great number of grammatical frameworks might agree upon the idea that analyses for the passive postulate more structure or additional steps to be taken in comparison to the ‘unmarked’ active case, i. e. viz. the assumed basic/‘canonical’ argument order projected from the verb’s lexicon entry. In addition, passive sentences differ from their respective active counterpart in that they contain an auxiliary phrase, as well as an (optional) prepositional phrase in order to roughly convey the same propositional meaning as an active sentence. Ceteris paribus, this additional structure should at least in principle lead to greater processing cost in terms of storage (see e. g. Culicover & Jackendoff, 2005). In addition, it might be argued that the alteration of the verb’s argument structure viz. the ‘canonical’ case in either syntax or the lexicon takes additional processing time.²⁰

¹⁸It is beyond the scope of this work to discuss why a particular argument structure representation which projects the objects to the right of the verb in English (and arguably to the left of the verb, but to the right of the subject in German) is to be assumed the ‘basic’, or canonical case (cf. Gorrell, 1995, 2000; also see Kayne, 1994 for a position that goes even beyond that; but see Haider, 2000 and S. Müller, 2010 for arguments why Kayne’s position might be questionable). Such language-specific differences have to be learned from the input. If the ‘canonical’ order of arguments in a lexicon entry is learned, and unless other external evidence exists to distinguish them, formal accounts like the one proposed by Drai and Grodzinsky (2006) are in fact indistinguishable from a frequency explanation (i. e. in essence chance).

¹⁹Compare Section 3.4.2 on the processing of relative clauses for arguments about the influence of information structure/discourse status, which might be applicable to processing for the active/passive diathesis as well.

²⁰Cf. however Shapiro (2004); Shapiro, Zurif and Grimshaw (1987), according to whom the
3. Sentence processing-for-production

3.4.2. Relative clauses

Relative clause structures have featured as a subject of investigation in several psycholinguistic studies. A number of different properties have been considered as possible reasons for higher difficulty when processing sentences with an embedded relative clause. One of them is the fact that relative clause structures involve embedding, for instance in the case of so-called centre-embedded sentences. Other work focused on the difference between subject and object relative clause structures.

F. Ferreira (1991) tested utterance latencies for structures of different complexity; her materials included relative clauses, the production of which was compared to the production of sentences which did not contain embedded structure. Ferreira’s results indicated that the additional structure (counted in terms of nodes in a tree graph; compare Yngve, 1960) results in a greater amount of time necessary for preparing utterances. This is corroborated by studies on sentence processing-for-comprehension, which also indicate that sentences containing embedded structure are more difficult to process. Gibson proposed a memory cost account (Gibson, 1998, 2000), according to which the storage and integration cost for new discourse referents are at the heart of processing difficulties. Gordon et al. (2001) point to a possible effect of interference between two potential role-fillers which are of the same type. Finally, Lewis (1996) showed how the processing limitations on center embedding might be explained to fall from general principles of working memory; also see his work with Vasishth, e. g. Lewis and Vasishth (2005). Because they explicitly frame their work in a comprehensive theory for cognition, especially Lewis and Vasishth (2005) can be assumed to model on the ‘algorithmic’ level in Marr’s sense. In comparison, is less clear where to locate the models by Gibson or Gordon and colleagues, since many aspects of the processing mechanism remain unspecified.

However, a problem with research on embedding per se is that the difficulty of structures with embedded clauses can be argued to be higher on both syntactic as well as semantic grounds: The conceptual structure of a

---

amount of possible argument structures associated with a verb’s lexicon entry contributes to processing difficulty, as the correct argument structure has to be selected for a given context. Crucially, the complexity of syntactic subcategorisation information (here: the amount of possible subcategorisation frames) is not reflected in processing time measures.
sentences containing a dependent clause is more complex, since the embedding relation must be encoded propositionally in some sort of conceptual (semantic) model; cf. e.g. Bloom et al. (1980); Schleppegrell (1992); Yuasa and Sadock (2002). In the context of the research proposed here, it is unclear what processing cost might be involved with the generation and maintenance of a conceptual model.

Because the two sentence types primarily differ in terms of structure, the difference between sentences containing subject and object relative clauses has been researched in studies on both production as well as comprehension. Evidence for differences in processing between subject and object relative clauses comes from several sub-fields of psycholinguistics, including research on language acquisition. In addition, in a typical acquisition path children remarkably often show difficulties with ‘non-canonical’ structures such as object relative clauses (Friedmann et al., 2009; Friedmann & Szternman, 2006; Hamann & Tuller, 2011).

Theoretically, the difference between SR and OR in English can be linked to word-order differences like in the active/passive diathesis or other alternations between SOV/SVO word order. The question is how to model apparent processing differences with the means of a grammar theory. Depending on the particular analysis, the amount of movement operations or dependency relations is identical for sentences containing a SR or an OR clause; so this factor might not serve as an explanation. However, if SOV is considered the ‘basic’ word order in German (Bierwisch, 1963), structures containing SR clauses should involve a shorter movement or dependency than sentences containing an OR clause. But it remains unclear what the sheer distance of a dependency relation should amount to in terms of processing cost, unless auxiliary assumptions like for instance processing time (by means of a ‘decay’ concept) or necessary memory space (cf. for instance Gibson, 1998, 2000) is invoked.

With respect to the possible function that the difference in word order between subject and object relative clauses might have, some authors have suggested an effect of discourse or information structure. In fact, it appears that factors pertaining to the discourse model influence both the acceptability of structures (Weskott, 2002), as well as processing speed in comprehension (Mak et al., 2008). In his 2002 dissertation, Weskott proposes that the perception of ‘markedness’ of a particular structure depends on the contextual restrictions imposed by information structure given a particular
discourse representation. According to the author, the information structure of a sentence forms a directive for the parser to search conceptual space in a particular way, in order to align the information given by the sentence with the current discourse model. This implies that a sentence will be regarded as acceptable, if the context provides sufficient restricting information to interpret a sentence in a ‘marked’ word order. This pertains especially to object-initial structures (OVS), given no special prosodic topic or focus marking (this line of argumentation could perhaps be extended to hold for other types of non-canonical structure as well, including the passive). If applied to sentence production, we might say that a ‘marked’ word order reflects a particular discourse or information structure, which differs from the canonical one, in which the Topic is mapped to the subject role/position.

A somewhat different factor that has been suggested as an explanation for apparent processing differences between sentences containing SR or OR respectively is similarity-based interference. Similarity-based effects of semantic features like animacy have been shown earlier in the literature by for instance Fanselow, Schlesewsky, Vogel and Weskott (2011); Lewis (1996); Traxler, Williams, Blozis and Morris (2005). A paper by Friedmann et al. (2009) investigates children’s perception of subject and object relative clauses in Hebrew and the authors argue that similarity-effects can arise on the basis of syntactic features as well. Crucially, their results indicate that processing of object relative clauses is not generally more difficult than the processing of subject relative clause structure. It has also been shown that in principle (healthy) adults have no problem to process object relative sentences (Belletti & Chesi, 2011). Both Belletti and Chesi (2011) and Friedmann et al. (2009) interpret their findings in terms of Rizzi’s principle of relativised minimality (RM).²¹ Processing difficulty is explained with recourse to “operative syntactic memory” (Friedmann et al., 2009: 84), and by assuming that children lack memory capacity to perform detailed enough comparisons between potential role fillers.

While the notion of “operative syntactic memory” remains unspecified in Friedmann et al. (2009), the author’s use of the factor similarity between ele-

²¹See Chapter 2, Section 2.2.1; repeated here for convenience:
In a constellation X << Z << Y the establishing of a syntactic dependency between Y and x is blocked if there is an intervening element Z that is structurally similar to Y, i.e. that is “of the same structure type” (cf. Rizzi, 2004).
3.4. Processing of different phenomena

ments in the structure being processed is quite interesting. The assumption is in fact very plausible in the light of what is assumed about processing in other cognitive domains. Especially if we take into account that sentences are processed under very narrow time constraints (Kolk, 1995; Levelt, 1989), the assumption that ‘similar’ elements might compete for being processed at a certain point in time (e. g. if a particular syntactic position or ‘slot’ is to be filled; cf. Kempen & Hoenkamp, 1987) appears quite plausible: The more ‘similar’ two parts of a data structure both activated at the same point in time or temporally close points in time, the more difficult it gets to either select the correct part of the data structure or inhibit the incorrect candidate. This kind of competition has in fact been at the heart of psycholinguistic or psychological explanations for priming effects for quite some time (e. g. see Glaser & Düngelhoff, 1984), although only recently these insights have been applied to sentence processing (Lewis, 1996).

During production, the syntactic marking of an information structural difference might require additional processing time to overcome a possible source of interference in terms of a locality violation. Syntactic interference effects might directly relate to properties of the processing mechanism, given that interference effects seem to occur in different cognitive domains (I will return to this idea in the discussion in Chapter 8). In addition, the problem attains a computational dimension in Marr’s sense, because of the necessity to substantiate the concept of ‘similarity’ (Ennis, 2007).

3.4.3. Agreement

A number of factors have been identified that seem to influence the generation of agreement in sentence production, including the grammatical relation between two elements, as well as semantic or morphological factors (Eberhard, Cutting & Bock, 2005: 531). The production of agreement, especially between subject and verb, has proven a highly interesting test case for language production research. Given the abstract function of the agree-

²²Compare Section 2.2.1 in Chapter 2 on the concept of similarity.
²³Similarity-based interference effects have been suggested as very general psychological processing principles (or ‘laws’) over a century ago in Gestalt psychology (law of proximity and the law of similarity); very much like many grammatical principles, however, these laws do not specify a processing mechanism but merely form a concise (although not formalised) description of observable phenomena.
ment relation, its establishment can be seen as a prototypical candidate for an unconscious, highly automatic processing step during grammatical encoding.

In Levelt’s language production theory, agreement is assumed to be computed on the fly during incremental sentence generation (Levelt, 1989; Levelt et al., 1999). It is generally assumed that the verb is specified for number on the lemma level, either indirectly through spreading of activation from the controller noun’s number specification (cf. Bock & Eberhard, 1993; Bock, Eberhard, Cutting, Meyer & Schriefers, 2001; Eberhard, 1997; Eberhard et al., 2005), or in addition to that based on conceptual or ‘notional’ information (cf. Vigliocco et al., 1996). Morphophonological processes retrieve or assemble the appropriate verb form based on the specified lemma. (Spoken) language production is a process that happens under narrow time-constraints (e. g. Van Turennout, Hagoort & Brown, 1998), and in fact the synchronization or fine-tuning necessary is remarkable: on the different level of phonemes, morphemes and words individual elements have to be produced in the right temporal order (F. Ferreira & Engelhardt, 2006; Kolk, 1995; Levelt, 1989).

The formulation stage, especially grammatical encoding, has been argued to operate without obvious resource limitations: “There appears to be a consensus that the core linguistic processes involved in speaking and listening do not pose strong capacity demands, in line with the lay view that speaking and listening are easy” (A. S. Meyer et al., 2007: xi). However, as mentioned in the introduction, there is an ongoing discussion about this issue. Some experimental evidence calls a strict assumption of modularity or automaticity into question, see for instance Garrod and Pickering (2007) for discussion. Evidence from experimental studies by Hartsuiker and Barkhuysen (2006) on agreement production and by V. S. Ferreira and Pashler (2002) on picture naming indicate that there might be processing load or effort incurred by the steps involved in formulation (also see Bock, Dell, Garnsey, Kramer & Kubose, 2007; Bock et al., 2003). Studies by Fayol et al. (1994); Hartsuiker and Barkhuysen (2006) have also called into question the notion of complete ‘encapsulation’, or of relative automaticity of

²⁴This holds for different modalities as well; a speaker of sign language for instance has to synchronise temporal order with the shape of signs and with the spatial layout of the signing space.
agreement generation during production, i.e. the question whether processing resources will be used for these processing steps. What is more, according to V. S. Ferreira and Firato (2002), the mechanisms that inform our language production system allow for syntactic variability as a result of the peculiarities (and shortcomings) of the processing system, for instance the availability of lexical elements. In particular, the unavailability of a particular lexical element can trigger structural revisions during language production (Bock, 1987; Levelt & Maassen, 1981).

Difficulties or errors can occur during processing, when a processing step runs out of memory space, processing time, or energy units (Hartsuiker & Kolk, 1998; Salthouse, 1988). In his account of agrammatic performance in speaking, Kolk (1995) for instance suggests that when running out of time, processing steps get ‘out of sync’ with subsequent or prior levels, and the data flow through different sub-systems is disrupted.

One crucial kind of evidence that has been informing different models of the on-line processing for language production in general and agreement production in particular has been data about so-called attraction errors. Quite a number of studies have exploited the attraction effect in order to investigate the processing of syntactic information in both spoken (Bock & Miller, 1991; Schriefers & van Kampen, 1993; Vigliocco & Nicol, 1998) and written language production (Fayol et al., 1994; Hemforth & Konieczny, 2003; Hölscher & Hemforth, 2000), as well as in written language comprehension (Pearlmutter, Garnsey & Bock, 1999; Wagers, Lau & Phillips, 2009). Different factors and conditions have been identified that proliferate the occurrence of number agreement errors as in example 15, where the number marking on the verb seems to be ‘attracted’ by an element other than the actual controller of verb agreement (cf. Section 2.2.3 for representational aspects).

(15) “However, it is only the meaning of the words which *have changed, not the grammatical structure of the language.”

Existing models for the establishment of agreement have focused different factors that exert an influence during the sentence production process, among them linear and hierarchical distance and semantic factors like conceptual number or distributivity (e.g. Hartsuiker & Barkhuysen, 2006;
Vigliocco et al., 1996; also see Chapter 2 and Chapter 7). As numerous studies have shown, the interference from local attractors is also subject to a number of other constraints, which influence when, and how many attraction errors occur. Bock and Miller (1991) tested the effects of animacy of the head and local noun phrase referent and found an increase in the amount of attraction errors when both referents were animate. Thornton and MacDonald (2003) interpret their experimental findings in terms of a probabilistic model that allows the parallel interaction of multiple constraints on sentence production, including the plausibility or ‘strength’ of the relation between a verb and potential controllers.²⁵

Most psycholinguistic models of agreement production concur that at the point of the verb lemma selection, the grammatical number feature of the verb has to be specified in so-called diacritic features, in order to retrieve (or generate) the correct word form. In Levelt’s original theory of grammatical encoding, building on Kempen and Hoekamp’s incremental procedural grammar (Kempen & Hoekamp, 1987), this feature specification is conceived of as a set of functions that inspect the conceptual message for number information of a (subject) noun and set the according value as the noun lemma’s diacritic number feature.

The marking and morphing model by Bock et al. (2001) and Eberhard et al. (2005) developed this concept further and specifies the inspection, or ‘valuation’ of conceptual structure in a similar fashion, yielding a notional number specification. The conceptual or notional number information is then used in parallel for lemma selection and marking, that is the specification of number on an abstract syntactic level. Subsequent positional processing or structural integration phase, morphing operations assemble morphological structure from the lexicon to create phrase structure. While doing so, possible conflicts between lexically specified number information and the specification inherited from the functional marking stage have to be

²⁵ Eberhard et al. (2005) provide a different explanation that maintains the relative independence of structure generation processes. They claim that both the results of Bock and Miller’s animacy manipulation, as well as Thornton and MacDonald’s study on plausibility can by subject’s difficulty to correctly identify the subject on the functional processing level. This should lead to errors that ‘mimic attraction’ (Eberhard et al., 2005: 555), but have their origin in a different mechanism than attraction (see also recent findings by Deutsch & Dank, 2009 for Hebrew). Based on self-paced reading data, Lau, Wagers, Stroud and Phillips (2008) argue that in comprehension, number retrieval on the verb pertains only to syntactic number features, rather than conceptual features.
resolved and agreement has to be established by transmitting number features from the subject noun to the verb. Eberhard et al. (2005) then suggest two points at which number information can be set incorrectly, a) during marking and b) if integration and reconciliation of lexical and structural information goes awry. With regards to the special case of attraction errors, the basic mechanism assumed is that number features of the interfering local (‘attractor’) noun spread or ‘percolate’ upwards though the network representing syntactic structure and in some cases overwrite the featural information coming from the actual controller. The verb lemma then ‘inherits’ the wrong number feature and a wrongly inflected word form is retrieved or generated.

The working memory retrieval model (Badecker & Kuminiak, 2007; Wagers et al., 2009) specifies a different mechanism for both correct as well as incorrect marking of agreement, based on work by Lewis and Vasishth (2005). Lewis and Vasishth (2005) recently formalised and implemented a model for real-time sentence comprehension in the ACT-R cognitive architecture (Adaptive Control of Thought-Rational; see for instance Anderson, 2007; Anderson et al., 2004). They model syntactic structure in the form of feature bundles or ‘chunks’ in memory for each maximal projection, with hierarchical relations between these objects encoded as feature-value pairs. A set of ‘production rules’ implements procedural knowledge, including grammatical constraints. In addition, a number of general architectural considerations about human cognition, based on empirical work from cognitive psychology, constrain the interplay of procedural knowledge and chunks of data in working memory as modelled by the ACT-R architecture. For instance, the amount of information that can be processed at a given moment in working memory is assumed to be limited. Estimations range from 7 (plus/minus 2; Miller, 1956) over 4 (Cowan, 2001), down to 1 object or chunk of information that can be hold in the focus of attention for manipulation (McElree, 2001). As Lewis and Vasishth (2005) note, these restrictions require repeated reactivations of chunks during the integration of a sentence that is being parsed, for instance if dependencies need to be resolved (see also Lewis, Vasishth & Van Dyke, 2006). This reactivation is based on an associative memory retrieval mechanism, which is assumed to be subject to similarity-based interference (e. g. Gordon et al., 2001; Gordon, Hendrick & Levine, 2002; Lewis et al., 2006; McElree, Foraker & Dyer, 2003).

As the model is squarely placed in the context of a general theory for
cognition and tries to explicate the processing mechanism with unprecedented potential for psychological realism, we can clearly speak of a model for the ‘algorithmic’ level of explanation of sentence processing. What is more, while the specifications in the paper by Lewis and Vasishth (2005) model parsing, it is possible to extended the approach rather straightforwardly to language production. Badecker and Kuminiak (2007) employ the notion proposed by Lewis and Vasishth’s model about the necessary reactivation of chunks in working memory based on retrieval cues for a model of agreement error production. Since sentence generation proceeds in an incremental fashion, the authors argue, earlier parts of the structure that are not in the current focus of attention any more have to be re-retrieved and brought back into the working memory ‘focus’ for further processing (McElree, 2001; see also Berti, 2010; Cowan, 2001). Syntactic structure would then be computed ‘on-the-fly’ by procedures that dynamically retrieve chunks in memory and set the relevant retrieval cues for further, super- or subordinate structural elements (see Marcus, 2013, for a recent variant of this idea). Retrieval cues are (sub-)sets of formal features that allow to identify other constituents, including previously activated ones. Examples are syntactic category, (nominative) case, features representing structural positions or grammatical functions (Badecker & Kuminiak, 2007; Lewis & Vasishth, 2005; Wagers et al., 2009).²⁶ During agreement production the subject noun phrase as the correct agreement controller has to be reactivated at the verb site. This happens in order for the noun phrase’s

²⁶ In addition to that, and following up on an idea by Thornton and MacDonald (2003) I suggest that in principle all information (syntactic, semantic/conceptual, morphological/phonological) about previously processed parts of a sentence might be available to establish retrieval cues—see however Lau et al. (2008) for data that might speak against this stance. In any case, syntactic features might form particularly fast or effective cues, due to the relatively small number and (relatively) clear-cut boundaries of the categories they establish.

In the production process, in cases where dislocated elements are produced somewhere at the beginning of the sentence, ‘traces’ might be seen as points of high probability to reactivate a lexical element, based on whatever syntactic, semantic or phonetic cues remain available at that point: see Gibson and Warren (2004) for an interpretation of traces as ‘memory buffers’, as critical points of reconstructing/reactivating possible referents in an incremental left-to-right construction process. On a formal level, this notion can be represented either by notions of slash categories, or by movement traces (however, compare Pickering & Barry, 1991, for an approach to reconstruction without empty categories).
number specification to be accessed, so that a correctly inflected verb form can be retrieved. According to the working memory retrieval accounts proposed by Badecker and Kuminiak (2007) and Wagers et al. (2009), agreement errors occur when there is interference from another, equally likely candidate that is sufficiently similar to the actual agreement controller.²⁷ If the wrong candidate is taken to be the agreement controller, a wrong cue is set for retrieval of the inflected verb form, which will result in an error.

As mentioned previously in Chapter 2, work on the so-called “plural markedness effect” has suggested that attractors marked for plural exert stronger attraction than singular attractors. This has lead for instance Eberhard (1997) to assume that agreement production is sensitive to morphological complexity, i.e. aspects of the organisation of the mental lexicon. Eberhard’s explanation of the effect pivots around the assumption that number marking is based on a privative or unary feature specification, and that the featureless, ‘unmarked’, or default number is singular. A ‘default’ marking or ‘singular bias’ has been suggested earlier by Hemforth and Konieczny (2003) as an explanation of their results from a written sentence completion study; also see Franck et al. (2004) for a similar argument based on acquisition data.

The explanation of a ‘default’ status for a particular number has been linked to the lexical frequency of an element (Franck et al., 2004). Crucially, there is ample evidence that the frequency of a particular inflected word form (or of a particular morpheme) plays a major role in word form access for both perception and production and can result in processing time differences: Word forms with a higher token frequency have been shown to be more easily or rapidly accessed (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965; Stemberger, 1984, 1985).

But the story is not quite that simple, since studies on morphological processing have shown that word form or token frequency alone seems to be

²⁷Notably, the formal ‘intervention’ account for describing agreement errors by Friedmann et al. (2009), based on Rizzi’s concept of relativised minimality (Rizzi, 1990, 2004) and Starke’s assumptions about featural similarity (Starke, 2001) is very much akin to the processing model by Lewis and Vasishth (2005). The ‘intervention’ concept is not only general enough so that it can be considered compatible with different grammatical frameworks and formalisms (Franck et al., 2006), but is also a good example for how a general cognitive principle operating on the syntactic level can help to explain linguistic phenomena (cf. Rizzi, 2004, p. 1).
an insufficient predictor of processing speed. The work by Kostić (1991); Kostić and Katz (1987) on Serbo-Croatian and related work has shown that in addition to frequency, the paradigmatic ordering of inflected word forms and syneretism within the paradigm plays an important role as well. As Kostić (1991) was able to show, the relative frequency per function in a paradigm acts as a highly precise, numerical predictor of naming latencies, for instance in lexical decision experiments. While frequency is negatively correlated with latency, the amount of syncretism within a paradigm appears to increase its complexity and hence decision RT (Milin et al., 2009; Van Ewijk & Avrutin, 2010). The observed reaction time patterns have been explained as the result of an interaction between frequency and distributional properties within a paradigm. This interaction can be numerically captured as the entropy, or informational content of a particular form, as a result of the functional transparency of a form and/or paradigm (Kostić, 1991, also see Baayen et al., 2007; Milin et al., 2009; Van Ewijk & Avrutin, 2010).²⁸ The smaller the entropy of a particular inflectional paradigm, the easier it will be to retrieve elements from this paradigm (Moscoso Del Prado Martín, Kostić & Baayen, 2004). With respect to the plural markedness effect, this might mean that a verb form marked for plural has a higher relative probability for ‘winning’ in a competition for activation, if there is any (even incorrect) evidence for a plural marking somewhere earlier in the sentence.

Overall, agreement (including agreement errors) appears to provide important insights about the architecture of the language production system, and the factors relevant in the computation of grammatical relations like subject-verb agreement. Observations about agreement errors might touch upon the question how informationally encapsulated the formulation stage actually is in general, and grammatical encoding processes in particular, i.e. whether these processes are dependent on (various amounts of) processing resources. In Chapter 4, I will return to this issue and present some conceptual assumptions that might allow us to further qualify the notion of automaticity of cognitive processes.

²⁸The operationalisation of entropy differs slightly between authors, as the studies cited ask different questions and apply different frames of reference for the calculation of relative frequencies. Kostić’s concept of entropy is based on the mathematical information theory by Shannon (1948a, 1948b), who established entropy as a measure for the amount of information in a message.
3.5. Summary

As stated at the beginning of this chapter, a fully-fledged model for the production of ‘complex’ or ‘difficult’ sentences under noise is beyond the scope of this thesis, since the focus of the work presented here is on the experimental exploration of the phenomenon of speaking under noise. Nevertheless, this chapter aimed to sketch some basic processing assumptions necessary to frame the empirical data described in part II of this work.

An important question that has to be answered with respect to a modelling of linguistic behaviour is the extent and scope of a model or theory. The experiments presented in this thesis deal with an empirical phenomenon that can be best construed as part of what is customarily considered ‘performance’: the influence of situational acoustic settings on the speaking process (cf. Levelt, 1974/2008). According to the three-tier hypothesis postulated by Marr (1982/2010), a performance phenomenon warrants a model for language processing on the level of the algorithm or mechanism. In the context of the ‘derivational theory of complexity’, a ‘realistic’ or naturalistic interpretation of a (variant of) formal grammar theory has been suggested as a possible blueprint for a processing mechanism. However, as the discussion about the reification of theoretical constructs has shown, this interpretation runs into severe epistemological problems. Many psycholinguists have therefore preferred to work with psychologically more plausible mechanistic models that are less strictly adherent to formal accounts. Naturally, we have to remain humble about the degree of empirical coverage to be expected from a mechanistic model. As for instance Phillips (2012) argues, it appears that current processing models are not yet detailed enough to capture the same amount of data as a more abstract grammar theory. What is more, I follow Marr’s argumentation that (at least for the moment) no unified, reductionist explanation of cognitive processes is foreseeable or feasible on an algorithmic/mechanistic level. Therefore both structural or computational issues, as well as processing factors pertaining to the algorithmic/mechanism level have to be taken into account. Finally, this humbleness should not serve as an excuse to reify theoretical concepts in psychologically pre-theoretic mechanisms.

From this (non-reductionist) dual viewpoint, I have tried to identify a number of possible candidate factors from earlier observations about sentence processing for both production and perception, which might serve
as (parts of) an explanation for why the processing of a particular kind of structure is ‘difficult’. First of all, the amount of structure comes to mind, and this factor has indeed been an considered important constraint on the effectiveness of processing. It could be argued that additional structure will incur additional processing cost in terms of memory space, as well as processing time. What is more, the narrow time-constraints under which the generation of sentence structure is carried out require the mechanisms to be exquisitely fine-tuned, in order to prevent errors. In addition, a monitoring loop (or multiple loops) serves to allow for error correction ‘on the fly’. When processing goes wrong, as in the case of agreement errors, the frequency and informational content (or entropy) of lexical elements might be at stake, that need to be distinguished at a given point in time. Authors like Lewis and Vasishth (2005) have argued that such problems are caused by similarity-based interference effects. Interestingly, interference effects reoccur throughout the arguments for processing difficulty related to the structures assessed in this study, for instance in the case of word-order differences in relative clauses. The processing of these structures too has been argued to be subject to linguistic interference phenomena.

The research question pursued in this thesis goes beyond this kind of ‘domain-internal’ interference, and in the following chapter I will discuss possible models for interference between external noise and the processing of sentence structure. In order to aim for an explanation of the interaction between language processing and momentary, external conditions, a model for a mechanism must provide theoretical and empirical ‘interfaces’ with a model for the effects of noise on cognition. A basic desideratum for the model is that the constructs it uses for e. g. human working memory are compatible with other psycholinguistic and more general psychological theorising and empirically sufficiently substantial.
4. Psychological aspects of speaking in noise

4.1. Introduction

As mentioned in the previous chapter, many a psycholinguistic metric or model for the processing of ‘difficult’ sentences uses constructs that aim to capture ‘capacity’ or something like the ‘cost’ of processing of a particular kind of sentence. Notions of ‘capacity’ or ‘resources’ have been playing an important role in cognitive psychology for research on other areas of cognition. Since these terms might refer to a number of different concepts, the question is how these notions can be substantiated in a psychologically plausible fashion. This chapter sets out to briefly sketch some of those details of the psychological theory necessary to research the phenomenon of using language under acoustically difficult circumstances.

The first section will deal with some earlier observations about differences between (groups of) subjects which have been argued to affect performance in linguistic tasks, namely age differences as well as differences in hearing acuity. For the research purpose of the work presented here, hearing acuity forms an important factor since the effects of external noise can be assumed to interact with the hearing status of a speaker experiencing noise. Both age and hearing acuity can be considered properties of an individual which change at a relatively broad time-scale, hence subjects can be grouped according to these properties in experimental designs.¹ The second section will describe previous results about the effects of noise on cognitive tasks.² For quite some time there have been studies about the ef-

¹I will refer to differences between subjects in terms of age and hearing status as inter-individual differences, since for the relatively short time period during an experiment each subject can be assigned a comparatively constant value on both scales.
²Under the controlled experimental conditions employed for the studies reported here, any detrimental effect of noise on performance can be traced back to the extrinsic cir-
4. Psychological aspects of speaking in noise

Effects different acoustic settings might have on human cognition, including our abilities to use language. Especially since psychologists and the general public alike have become aware of an increase in everyday noise at work places, or anywhere related to traffic (up to the point of considering it a type of ‘pollution’) there has been a focus on gauging possible detrimental effects of noise (e. g. Hygge, Evans & Bullinger, 2002; also see Smith, 1989, for a review of noise effects on performance in general). The chapter will close with a discussion of concepts and hypotheses from cognitive psychology on noise-induced interference effects. I consider these concepts highly useful for an explanation of the possible interaction between language processing and factors like noise, age, and (working) memory capacity.

4.2. Individual differences in processing-for-production

4.2.1. Effects of hearing impairment

Experiencing a situation with external noise is qualitatively different from the sensory experience of people with hearing impairment. Nevertheless, both experiences share the property of imposing limits on the auditory sensory input during the speaking process. What is more, noisy environments have been reported to impose particular difficulties to individuals with hearing impairment. For language comprehension, a number of studies have compared the simple and interactive effects of comprehending language under noise and/or with hearing impairment (Larsby, Hallgren, Lyxell & Arlinger, 2005; Smoorenburg, 1992; Uslar et al., 2010, 2013). The reduced spectral and temporal resolution typically found in listeners with sensorineural hearing impairments adds up to the difficulty imposed by noise masking alone (Moore, 1985; Wingfield, Tun & McCoy, 2005).

With respect to the effects of hearing impairment on the speaking process, it is important to distinguish between the developmental, acquisition-related effects of hearing impairment and the more immediate situational effects of the signal degradation on on-line processing. A way to disentangle the two kinds of effect is to test populations with different ages...
of onset of the hearing impairment. Unlike children with *congenital* or very early hearing loss, subjects with *post-lingual* hearing loss acquired the impairment after important stages of language acquisition had been completed. Their performance on linguistic tasks then should not reflect effects of degraded input on the acquisition process, but rather momentary difficulties caused by the signal degradation. Spoken language by children with (congenital) hearing impairment has been analysed in a number of studies on language acquisition, see for instance Brannon (1968); Friedmann and Szterman (2006); Geers and Moog (1978). Work focusing on syntactic abilities has shown that hearing impairment or hearing loss that is not treated until the age of about 8 months (Friedmann & Szterman, 2006) can lead to different courses in the language acquisition process (also see for instance Delage & Tuller, 2010). However, a more detailed discussion of the effects of congenital hearing impairment is beyond the scope of the work presented here, and in the empirical part I will focus on the effects of post-lingual hearing impairment.

At present, it is unclear whether a post-lingual hearing impairment has an effect on the grammatical complexity of structures produced. According to authors like Leder et al. (1987) or Pratt (2004), *presby(a)cusis* or age-related hearing loss, the prototypical case of a post-lingual impairment, is sometimes accompanied by changes in speaking rate or precision of articulation. These differences, however, only seem to surface when the hearing loss is severe, i.e. complete or near-complete. To our knowledge, there are few if any references to the effects of post-lingual hearing impairment on more central, grammatical aspects of language production, let alone (quasi-)experimental studies testing this. Goehl and Kaufman (1984) for instance posit that in speakers with adventitiously acquired hearing loss “knowledge of the syntax, semantics, and phonology of the native language will remain intact” (p. 59; also see Pratt, 2004). Since a systematic investigation of the effects of hearing impairment on the syntactic aspects of sentence production seems to be lacking, study 2 has been designed to provide further empirical evidence about this question (see Chapter 6).

**4.2.2. Effects of ageing**

Especially research on speakers with post-lingual hearing impairment has to deal with the fact that the prevalence of this kind of disorder is highest in
4. Psychological aspects of speaking in noise

the elderly, in the form of presbyacusis or age-related hearing loss (Dubno & Mills, 2004; Wingfield et al., 2003). Hence it has to be borne in mind that in experiments on post-lingual hearing impairment the factor hearing status can interact with age.

It forms a common observation that with increasing age, cognitive abilities deteriorate (Salthouse, 2004). Intuitive observations to that end have been investigated systematically in gerontology, which essentially corroborated the subjective impression with cognitive measures, see for instance Salthouse (2001) for a review. Cognitive effects of ageing have been described as pertaining to processing speed or efficiency, assuming a slowing down of cognitive processes with increasing age (Salthouse, 1988, 2000, 2004). An alternative but not necessarily incompatible view relates performance reductions in older age to a reduced ability to overcome cognitive interference (e.g. Connelly, Hasher & Zacks, 1991).

With respect to linguistic comprehension abilities, Wingfield et al. (2006) for instance found an interaction between effects of ageing and hearing impairment on language comprehension. Their results suggest that if multiple factors influence language processing in a detrimental way, speakers-listeners reach a ceiling (or floor) level after which performance declines in a what Wingfield et al. call “multiplicative” fashion (also see Section 4.4.1).

Given these findings, it forms an important question how ageing might affect speaking, especially the generation of syntactic structure. Susan Kemper and co-workers investigated the effects of ageing on language production and found effects of reduced syntactic complexity in naturalistic corpora of written language (diary entries, Kemper, 1987), as well as in (quasi-)experimental studies testing both written as well as spoken language production (Kemper, 1986; Kemper et al., 1989). Kemper and Sumner (2001) not only found a significant effect of age on grammatical complexity of utterances, but also significant correlations between language production and measures of cognitive ability like reading or digit span (cf. also Zekveld, George, Kramer, Goverts & Houtgast, 2007 and Van der Linden et al., 1999). For language production, this assumption was tested by Cheung and Kemper (1992) using different metrics of linguistic or processing complexity in language production. The authors explain their results with recourse to the notion of ‘working memory’ (see Section 4.4.2) and conclude that “verbal ability and working memory are correlated factors that change with advancing age and determine how speakers’ sentence length,
the amount of embedding, and the type of embedding vary with advancing age.” (Cheung & Kemper, 1992, p. 69).

Taken together these results show that certain structural factors can create greater processing difficulty in elderly speakers, and they indicate that cognitive (or more specifically working memory) capacity plays a role in this relationship. In particular, age and linguistic performance might be indirectly related, through cognitive factors (e.g., reduced processing speed or increased susceptibility to interference) that act as mediators (cf. also Rabbitt, 1991).

4.3. Moment-to-moment differences in performance

4.3.1. Effects of noise on language comprehension

The effects of noise on language comprehension are first of all obviously linked to the (energetic) masking properties of a noise signal, which renders parts of the ‘useful’ signal unusable. The addition of noise signal and speech will result in interference if not ‘drowning’ of the useful speech signal by the noise (B. A. Schneider, Li & Daneman, 2007). However, audiological and psychoacoustic research has emphasised the role of the acoustic characteristics of background noise created by multiple speakers talking (see Bronkhorst, 2000 for a review). For instance, the ‘buzz’ or ‘babble’ in situations with many concurrent speakers possesses particular acoustic characteristics that are different from ‘static’ (white or pink) noise like the humming of an A/C fan or the engine noise inside an aircraft. The temporal characteristics of a ‘babble’ noise signal can be described as showing broad-scale fluctuations in amplitude across time, creating an envelope of intensity peaks and troughs, which are partially responsible for creating the ‘babble’ impression. Figure 4.1 on page 76 illustrates this aspect of different noise types schematically.

In research on language comprehension, noises with different temporal fluctuation parameters have been shown to exert different effects on intelligibility. Temporally fluctuating (‘babble’) noise has a less detrimental effect on comprehension of language stimuli than constant noise of the same average intensity (Cooke, 2006; Cooke, Lu, Lu & Horaud, 2007; Wa-
4. Psychological aspects of speaking in noise

Figure 4.1: Schematic drawing of the broad-scale intensity fluctuation of different noise types. The fluctuation pattern of the actual sounds is random, the regular pattern shown here has been chosen for illustration purposes only.

gener, Brand & Kollmeier, 2006 and references therein). This difference has been linked to the availability of gaps or “dips” of relative silence in fluctuating noise, which allows for “glimpsing” (Cooke, 2006) at the to-be-comprehended stimulus. The benefit of these glimpses at some (however random) portions of an utterance appears to outweigh, at least on average, the potentially stronger masking of other parts of the ‘useful’ signal by intensity peaks present in the noise signal. Research by for instance Festen and Plomp (1990), Wagener et al. (2006) and Cooke (2006) on language comprehension compared the effect of different kinds of masking noise on speech intelligibility. In these studies, interfering sound signals with temporal fluctuation were compared to signals without, while keeping the average (root mean square, RMS) intensity at the same level. Necessarily, in order to reach the same average, the two types of sound will differ in their maximum amplitude, and the fluctuating noise will contain peaks of high intensity as well of troughs or “dips” of lower intensity compared to the mean. The results show that on average, listeners can benefit from the “dips” that exist in a masking sound with fluctuating intensity (see e. g.
4.3. Moment-to-moment differences in performance

Difficulties with language comprehension in noise seem at first sight rather straightforwardly a result of masking of the signal, as in this case the actual acoustic representation of a sentence or word is changed before it reaches the ear of the comprehender. However, processing-for-comprehension under adverse conditions with signal degradation appears particularly difficult when complex, non-canonical, or otherwise ‘unusual’ sentence structure is being processed. Besides the physical loss of information in the summed signal of noise and language (Cooke, 2006; Wagener et al., 2006), it has been assumed that cognitive factors play a role in the comprehension of language in noise. As part of the data in the signal is masked, a listener has to try to ‘restaurate’ information about phonemes or words, in order to be able to construct a complete linguistic representation for a word or a sentence. Experiments by Wingfield and colleagues yielded effects of comprehension difficulty that seem to not only bear on the necessary restauration of phonological information for individual words, but also on the difficulty of the syntactic structure that has to be processed during a comprehension task. Wingfield et al. (2003) for instance contrasted different types of sentence structure that had been perceptually degraded. They found that the processing difficulties created by time compression and by the linguistic factors do not simply add up, but interact: the greater the perceptual degradation, the more difficult it becomes for subjects to process complex sentence structures. This effect has also been replicated for different age groups and populations with hearing impairment, for whom comprehending language in acoustically challenging environments presents a particularly difficult task (Carroll & Ruigendijk, 2013; McCoy et al., 2005; Wingfield et al., 2006, 2003, 2005). As for an explanation of these phenom-

³Quite rightly, however, Bronkhorst (2000) objects that “[u]nfortunately, this effect cannot be modeled easily because it depends not only on the envelope [i.e. the intensity fluctuation, M. H.] of the interferer, but also on the properties of the target speech.” (p. 120). The individual elements (morphemes) of a sentence do not uniformly contribute to the interpretation of a sentence, but are highly structured: Some elements form stronger or more important cues for establishing for instance the thematic relations of a sentence than others. If the low intensity portions by chance occur at the same points in time as those morphemes indicating grammatical functions, this kind of information will be preserved. Unless we are able to control the signal-to-noise ratio (SNR) on a morpheme- or even phoneme-by-phoneme basis, predictive modelling of intelligibility as a function of noise type will necessarily remain rather coarse.
4. Psychological aspects of speaking in noise

Wingfield and others have resorted to ‘capacity’ notions. I will return to the cognitive aspects of processing language in noise in section 4.4.1.

4.3.2. Effects of noise on speaking

As I have illustrated in the introduction, it is intuitively obvious that the situational acoustic circumstances have an influence on us when we speak. So far, empirical work that has dealt with the effect of noise on speaking has been concerned with for instance acoustic or prosodic effects, speech rate and the amount of speech errors.

A quite extensive strand of research on language production under the influence of noise in fact comes from acoustics and phonetics, and is concerned with effects on articulation. Effects of noise on pitch and intensity have been described as early as 1911 by Étienne Lombard. This effect, often referred to as Lombard reflex or effect is marked by an increase in perceived articulation effort (Junqua, Fincke & Field, 1999). Effort, in this sense, can be broken down into more detailed perceptual measures, such as overall pitch and intensity, which correlate (non-linearly) with frequency and amplitude of a signal (cf. e.g. Junqua et al., 1999; Van Summers, Pisoni, Bernacki, Pedlow & Stokes, 1988; Waldstein, 1990). Van Summers et al. (1988) also report findings on fine spectral changes in individual phonemes that characterise noise-masked, or “lombardic” speech. A recent study by Lu and Cooke (2008) indicated that speech produced under noise is perceived as more intelligible (also see Dreher & O’Neill, 1957). In a series of experiments testing for ‘active’ components in acoustic changes by speakers the authors could show that depending on spectral characteristics of noise, speakers make characteristic adjustments to their speech. It is however unclear to what extent these ‘active’ changes reflect voluntary, i.e. attentionally controlled changes.

Crucially, according to Lu and Cooke there is only weak evidence for noise-induced modifications on other (categorial) levels. In terms of Levelt’s model of language production, the effects on the acoustic level should primarily pertain to the articulation stage. Thus the question remains how earlier levels of processing, especially those pertaining to linguistic structure might be affected.

Previous studies that looked at the effect of noise on speaking at a sentential level pointed to another noticeable effect on speakers, that of changes
4.3. Moment-to-moment differences in performance

in speech rate and utterance duration. In an early demonstration of the effect of noise on speaking, Hanley and Steer (1949) showed a significant reduction of speech rate in words per minute. These observations have been corroborated with more detailed analyses in newer work, e. g. by Leder et al. (1987) or Van Summers et al. (1988).

Postma and Noordanus (1996) used noise to interfere with processing during monitoring for speech errors. The authors tried to block the external feedback loop of speakers producing tongue-twisters by presenting them with white noise at a high intensity through headphones. Their results again pointed to effects mainly on the articulatory level, as the authors showed that noise led to a slowing down of articulation and to more phonological errors going unnoticed by the speakers. The authors account for their results based on the 'levels of integration' language production architecture (Garrett, 1980; Levelt, 1989), more specifically on the assumption of a monitoring system that checks language production on-the-fly for errors. Postma and Noordanus (1996) argue that monitoring through the external feedback link provides an additional channel to detect errors, which can be blocked or impaired in noise. However, the task employed relies on conscious search for errors, which makes it difficult to assess the effect of perceptual monitoring on automatic processing of morphosyntax under noise. For this reason it is difficult to say whether the authors’ conclusion can be generalised to errors that occur during the assembly of syntactic and/or morphological structure.

Finally, results by Kemper et al. (2003) indicated that difficult acoustic surroundings might indeed have an effect on earlier stages of utterance generation. The authors analysed several characteristics of spoken language production by different populations when a secondary task had to be performed alongside the (primary) speaking task. Crucially, their study included an experimental setting where subjects were presented with ‘cafeteria babble’ background noise and were asked to ignore it. The authors report differential effects of conditions with a secondary task on grammatical complexity, also contingent on the age group speakers belonged to: while older participants in their study relied on a reduction of the rate of speech, younger participants produced speech contained fewer embeddings. For their measurements, Kemper et al. (2003) operationalised grammatical complexity of utterances by calculating the mean number of clauses per utterance (MCU), and by determining the ‘D-Level’ score for each clause. D-
4. Psychological aspects of speaking in noise

Level refers to a compound score of grammatical complexity based on the age of acquisition for particular types of sentence structure (Rosenberg & Abbeduto, 1987).

4.3.3. The irrelevant sound effect (ISE)

As the results by Kemper et al. (2003) indicated, speakers can be taxed by situational difficulties and try to cope with them by resorting to ‘easier’ sentence structures. An obvious question in this respect is how simply listening to or ignoring speech or noise might lead to a reduction of performance.

A possible answer might be found in literature about the so-called ‘irrelevant speech effect’ or ‘irrelevant sound effect’. Several decades of empirical research about the irrelevant speech effect has converged upon the finding that concurrent speech has a detrimental effect on processing for tasks that involve verbal material. In typical ISE studies, experimental tasks like recall of word or digit lists were used as a dependent measure while presenting speech material of different kinds as distractors. If irrelevant speech is present, performance on a recall task for verbal material declines. Somewhat counterintuitively, the effect seems to be independent of intensity or content of concurrent speech.

Crucially, such effects are not limited to irrelevant speech. As for instance Jones and Macken (1993) as well as Klatte and Hellbrück (1993) have shown, series of intermittent tones or speech-free noise lead to performance decreases. What is more, Klatte, Kilcher and Hellbrück (1995) have shown that the negative effect on performance is substantially stronger with fluctuating, ‘babble-like’ noise than with constant noise. Hence, these results indicate that simply listening to, or ignoring certain kinds of noise can draw on cognitive resources necessary for other tasks. This has important implications for the question of automaticity. Before I turn to this issue, however, a discussion of what is actually meant by ‘resources’ is in order.

4.4. Cognitive resources and interference

A number of different observations about the influence of noise on language processing have been explained with recourse to concepts of (e. g. memory) ‘capacity’ or ‘processing resources’. In order to frame the debate
conceptually, it is assumed that processing requires some sort of resource in terms of ‘energy’, ‘time’ or (storage/buffer) ‘space’ (see Salthouse, 1988; compare for instance Pichora-Fuller et al., 1995; Rabbitt, 1991; Wingfield et al., 2003 for applications of these concepts, but see Ellen R. Stoltzfus, Lynn Hasher & Rose T. Zacks, 1996). More generally, the notion of limited resources has played a critical role in the discussion about individual differences in linguistic performance, or variance between subjects (cf. e.g. Just & Carpenter, 1992). The assumption that cognitive processes are bounded by capacity limits is central to the concept of interference by a secondary task load, e.g. in studies employing a dual task paradigm.

4.4.1. Dual-task performance

For quite some time the effects of doing two or more things at once have formed a central research focus for cognitive psychology in work on so-called dual task paradigms: Intuitively, tasks like walking, talking, eating, etc. seem rather straightforward and automatic, in that we usually do not need to devote much attention to them. However, for situations where many tasks with little demands co-occur, it has been observed that performance wanes (Heuer, 1996; Pashler, 1994).

With respect to language comprehension under noise, one of the key results of research by Wingfield and colleagues was that detrimental effects of age, hearing impairment and noise on language comprehension can interact (Wingfield et al., 2005). Wingfield and colleagues have framed the interaction of perceptual difficulty and cognitive effects on performance within Rabbitt’s effortfulness hypothesis (Rabbitt, 1968), also see McCoy et al. (2005); Pichora-Fuller et al. (1995); Rabbitt (1991); Wingfield et al. (2005). Already in 1968, Rabbitt discussed the effect of perceptual degradation and necessary signal restoration, and he proposed that signal loss due to hearing impairment might be seen as a imposing a secondary task load on the subject. In 1991, he wrote: “it is useful to use the metaphor that mild deafness, in effect, imposes an additional ‘secondary task’ which must be carried out whenever speech has to be understood, conversations managed or monologues interpreted.” (Rabbitt, 1991, p. 175). The claim by for instance Wingfield et al. (2006) is that a degradation of stimuli, through noise masking and/or hearing impairment requires greater effort during perceptual processing in order to ‘repair’ the acoustic signal and extract as much
Together with the assumption that resources required for processing on the perceptual level and further “downstream” (Wingfield et al., 2005) are limited and shared across cognitive domains, the following view on language comprehension under adverse conditions emerges: The degraded stimulus has to be identified at greater processing cost, incurred by for instance the necessary integration of different sources of context information (phonetic, syntactic, semantic, pragmatic; Pichora-Fuller, 2008). In order to muster the additional resources, processing space, time, or energy (Salthouse, 1988), or a combination thereof, are reallocated and will not be available any more for processes drawing on the same resource pool. A slightly different construal is that perceptual and other processing steps compete for resources on a ‘central bottleneck’, see for instance Wingfield et al. (2003). Concurrent language processing will have fewer resources available and will become more likely to fail or will slow down. Such a concept of ‘channel capacity’ is essentially a combination of the resource metaphors time and space, since it defines capacity in terms of the amount of data per time.

If cognitive processing is slowed down or has fewer resources available to begin with because of ageing (Salthouse, 2004), the additional burden of processing degraded stimuli will cause a kind of ‘resource floor’ effect, since resources will be taxed up to a level where no spare capacity is available to restore a degraded signal, and comprehension difficulty ensues (Rabitt, 1991; Wingfield et al., 2006, 2005). For Wingfield and colleagues this serves as an explanation for the interactive effects of signal degradation, age, hearing impairment and linguistic complexity that were reported by for instance Wingfield et al. (2006) and others: since the overall ability to overcome perceptual difficulties is reduced in older listeners, they will reach their resource limits earlier than younger speakers. Listeners with hearing impairment will be particularly challenged by adverse background noise conditions, since they suffer from additional signal distortion. While such an explanation seems plausible for the case of language comprehension, it is unclear how it could extend to the case of speaking under adverse circumstances.

Kemper et al. (2003) interpret their results on language production under noise as evidence for the assumption that simply listening to noise can exert a dual task effect, which in turn can influence sentence formulation.
4.4. Cognitive resources and interference

processes. What is more, the authors found significant correlations with different measures for working memory capacity constructs that had been argued to be relevant in language processing. Fayol et al. (1994) report evidence from written production for an effect of dual-task load on the likelihood of attraction errors. Their explanation is based on the assumption of a resource-constrained, attentional checking process that deteriorates in its ability to prevent the verb from inheriting a ‘percolated’, erroneously set number specification. Thus a central, mediating role in the explanation of dual-task effects on speaking seems to be played by the notion of limitations in processing or working memory capacity. Various authors like Gibson (1998), Kemper and Kemtes (1999) or Kemper and Sumner (2001), to mention only a few, have argued that individual capacity constraints form an important boundary condition for the processing of difficult types of sentence structure. Similar to the effects of ageing, Kemper et al. (2003) claim that difficult, noisy situations are taxing on processing or memory capacity and that speakers have to adjust for instance the amount of embedded clauses during speaking as a kind of ‘coping strategy’. These results not only indicate that ignoring irrelevant noise can generate a secondary task load, but also suggest that this load can impact the generation of linguistic structure during the grammatical encoding stage—contrary to what might have been expected from the assumption that processing on the formulation stage happens automatically or resource-free (cf. Section 3.2.1).

Researchers like for instance V. S. Ferreira and Pashler (2002) tried to experimentally distinguish between modular (or specialised) processing mechanisms and central (or shared) mechanisms used for language production. The study by V. S. Ferreira and Pashler (2002) tested the effect of a naming tasks with different linguistic constraints on a secondary, unrelated tone discrimination task. Their results were interpreted as evidence for a central bottleneck that influences early stages of word-production, but not later stages. However, the results by Kemper et al. (2003) suggest that the grammatical encoding stage might also be affected by a capacity reduction. The two apparently contradictory results might be reconciled by looking at further evidence from dual-task experiments. As Heuer (1996) notes, performance problems occur most clearly when two or more concurrent tasks are similar in some way or other. Bronkhorst (2000) remarks on language comprehension in noise that: “it is, indeed, probable that excess masking occurs when target and interfering voices are so alike that speech segrega-
4. Psychological aspects of speaking in noise

...tion becomes difficult” (p. 119). And Macken, Tremblay, Alford and Jones (1999) conclude the following from their work on the ISE: “More generally, where a memory task involves, for example, the processing of semantic information, then the semantic properties of interfering material will be important” (p. 323). Thus the assumption of a single resource pool that is shared across cognitive functions or domains has to be called into question.

4.4.2. Working memory

Not only in the context of dual-task paradigms the capacity question has been tightly linked to the concept of working memory, a system or ability that allows us to retrieve, manipulate and maintain information (Baddeley, 1986, 2003b; Nairne, 1990). Limitations of working memory are regularly used as an explanatory construct for individual differences in cognitive performance across domains, including linguistic performance (Just & Carpenter, 1992; Salthouse, 1988).

Given the significant differences in how competition for cognitive capacity is conceptualised for instance in studies on the joint effects of ageing, hearing impairment and noise on language comprehension, it is difficult to pinpoint the processing stage at which an interaction is supposed to occur, since this question largely depends on basic assumptions about the architecture of working memory and its available resources. To this day, the concept of working memory is the subject of ongoing discussion about both its general architecture, as well as how to measure its capacity (Berti, 2010; Conway et al., 2005).

Working memory models

The classical conceptualisation of working memory to be found in psychological literature and also widely cited in psycholinguistic work is Baddeley’s modular model (Baddeley, 1996, 2000). In his conception, working memory is a “workbench”, on which information is encoded, retrieved, and manipulated.⁴ Baddeley’s memory model consists of a main workbench component, the so called central executive, two modality-specific subsystems called

⁴In this respect, the model is different from the short term/long term memory perspective, which is rather concerned with timing and capacity aspects of storage (cf. Jonides et al., 2008).
visuo-spatial sketchpad and phonological loop, and an integrative episodic buffer component that is assumed to interface with long term storage. The role of the phonological loop is to maintain (auditory) representations, until they can be processed further by the central executive.

In the debate about the architecture of working memory, especially the aspect of overall modular in the conception by Baddeley has been called into question. Domain or modality specificity, that is processing a particular type of content does not necessarily require the assumption of functionally completely separate components or modules. Cowan’s model for human memory assumes that working memory is more of a state than a location, keeping information from long term storage or sensory input in an activated state ready for processing. A central role in this concept plays the notion of a capacity-limited focus of attention that can hold between one and four items (cf. Cowan, 2001; McElree, 2001; see Berti, 2010 for a recent review). According to Berti (2010), Cowan’s conception aligns well with models from neuropsychology that conceptualise working memory not as a functionally and anatomically unified structure, but rather as an emergent property of the interaction between individual processes for storing or altering information and attentional control (see for instance Postle, 2006).

**Verbal working memory**

In psycholinguistics, there has been a long-standing debate about the role of working memory in language processing. Several studies on sentence processing have been explaining complexity phenomena by recourse to differences in working memory capacity or ‘processing resources’ (see references in Section 3.4 on page 53; also see for instance Franck et al., 2010; Friedmann et al., 2009).

Different models have been proposed to explain how humans ‘work’ with data from different sources in order to achieve for instance communicative goals. Part of this debate has been the question whether one should assume a special component or set of resources necessary for processing language. Such a set of resources has been proposed by Just and Carpenter (1992), who called it verbal working memory (VWM; see also Daneman & Carpenter, 1980). The proposal of a VWM model by Just and Carpenter (1992) has been followed by a debate whether this component can indeed
be differentiated from other WM components, and whether it should be further subdivided into smaller resources for different linguistic processing steps (e.g. Caplan & Waters, 1999; Waters & Caplan, 1996)—or whether a symbolic processing architecture is a probable concept at all (cf. e.g. MacDonald & Christiansen, 2002). As a consequence of the differences in modelling a working memory system, the notion of capacity can differ quite substantially, as do the possible methods to measure it (see Conway et al., 2005 and Waters & Caplan, 2003 for discussion).

Just and Carpenter (1992) for instance assume a unitary, energetic processing resource, which is debited more or less strongly by different processing steps. The authors maintain that their model provides an explanation for the apparent modularity of syntactic processing which is according to the authors not based on an architectural separation between ‘encapsulated’ modules, but rather because resource limitations constrain possible interactions between syntactic and what they call “pragmatic” processing, pertaining for instance to animacy of the referents.

This architectural conception has been contested by Waters and Caplan (1996); also see Caplan and Waters (1999); Waters and Caplan (2003). The authors’ criticism is based on a range of neuropsychological evidence and due to concerns about the validity of the reading span task, the tool proposed by Daneman and Carpenter (1980) and Just and Carpenter (1992) to measure individual differences in verbal working memory span (see also Conway et al., 2005 and Garrod & Pickering, 2007 for discussion). In particular, they criticise the strong reliance on conscious memory retrieval and that the to-be-stored verbal material in the task is actually unrelated to the processing component of the task, unlike in sentence comprehension (cf. Waters & Caplan, 1996, p. 769). Waters and Caplan (1996) instead suggest a more differentiated verbal working memory, which consists of at least two different kinds of resources for verbally mediated tasks. One of these pertains to unconscious and highly automatised (obligatory) linguistic processing, including acoustic, phonological, syntactic, semantic

⁵Note that the concept of verbal working memory, as for instance explicated by Just and Carpenter (1992) cannot be identified with the phonological loop component of the model by Baddeley (1986); rather, Just and Carpenter (1992) state that their conception of a working memory for language “corresponds approximately to the part of the central executive [emphasis added] in Baddeley’s theory that deals with language comprehension.” (p. 123).
and pragmatic processing. A different set of resources is required for “conscious, controlled, and verbally mediated processes, such as the deliberate search through semantic memory for a piece of information, explicit reasoning, and other tasks” (Waters & Caplan, 1996, p. 770). Caplan and Waters (1999) further claim that the apparent automaticity or constraint-free working of core linguistic processes (including phonological, morphological, syntactic and semantic processing) is due to the high domain-specificity of these processes, with relatively few concurrent tasks competing for space or time on underlying cognitive procedures. This might be different for what the authors call “post-interpretive” processes, including attentional, conscious reasoning.

**Cognitive modularity**

The suggested fractionation of (verbal) working memory has severe implications for the question of how external noise, age and hearing impairment can interfere with language comprehension and language production. A possible explanation bears strongly on the question of cognitive modularity (Fodor, 1983, but also see Fodor, 2000). The question has for instance been raised by Caplan and Waters (1999). At first sight, their “separate language interpretation resource” (SLIR) hypothesis appears to support the notion of modularity, especially Fodor’s concept of informational encapsulation. But according to Caplan and Waters, their model does not require informational encapsulation. They state that (unconscious or implicit) language processing is the combination of mapping processes between phonological, syntactic, and semantic structure, that typically occurs over this kind of data and add: “Very few functions other than language interpretation [phon./syn./sem. structure processing; M. H.] compute any of the intermediate representations that are computed by the interpretation process, and none computes all the representations that are routinely computed by this process” (p. 93).

The same insight has been theoretically explicated on a more general level by Barrett and Kurzban (2006), who argue for a reconceptualisation of cognitive modularity in terms of a domain- or data-specific *functional*

---

*Note that Wingfield et al. (2006) posits that age-related processing difficulties very likely stem from the second, post-interpretative processing stage and do not involve automatic parsing processes.*
specialisation of processes. Specialised processes operate (possibly in parallel, see also Jackendoff, 2002, on this matter) only on particular types of data structure, that is only those that are compatible with the respective input criteria of any given process. Following this argument, the view taken here is that the “modular” makeup of a cognitive system like language as it has been viewed in the aftermath of Fodor’s conceptualisation can be best construed as epiphenomenal of a system in which individual operations on perceived or stored and retrieved data work together to achieve ‘higher-order’ functions (cf. Barrett & Kurzban, 2006; also see Berti, 2010 for a similar view on the concept of working memory). Partial overlap between any of these higher-order functions can occur through sharing of basic operations over particular data types. The distinction (implicitly or explicitly assumed in much work in psycholinguistics and linguistics) between linguistic processing and general cognitive processes rather turns out to be a matter of automaticity and domain-specificity of basic operations. In the following, I will show how explanations for the irrelevant sound effect have applied these concepts theoretically.

4.4.3. The changing-state hypothesis

The body of work about the irrelevant sound effect (ISE) is concerned with mainly one experimental task: serial recall. Also, many studies deal with the effects of concurrent speech on verbal recall, this way implicitly introducing an additional factor of linguistic processing into the model. Some studies have tried to capture this additional factor in more detail and claim that the semantics of words presented as irrelevant background stimuli exert an effect on serial recall (e.g. Beaman, 2004; Oswald, Tremblay & Jones, 2000). In this kind of study, however, the acoustic/perceptual effects and

---

⁷To make this conceptual line of argumentation less abstract: I consider a ‘higher order’ function to be something like “producing a sentence” or “recognising a sound pattern”. ‘Basic’ operations should be on the order of unification or merge, or even more general than that, like storing, retrieving and comparing of data chunks.

⁸A similar position has recently been advocated by Chomsky, cf. Chomsky (2004):

> domain-specificity is reduced to some special arrangement of elements that are not language-specific. . . . [I]t could turn out . . . that every component of the faculty of language is doing something else. And that what’s domain-specific is the organization of these various components. (p. 162)
4.4. Cognitive resources and interference

effects of higher order processing of language might be confounded. What is more, it might prove quite difficult to extrapolate the results based on a serial recall task to other, perhaps more complex cognitive functions like speaking. And since speaking crucially involves the generation of linguistically structured material rather than recalling ordered lists of otherwise unconnected words, an obvious, yet still open question is whether (and if so, how) noise has an impact on any of the various cognitive processes that allow us to produce language. More precisely, it is still unclear which steps or functions performed during natural language processing-for-production will be susceptible to similar effects of a dual-task load within a functionally specialised (verbal) working memory, and more importantly what the mechanism for the interference could look like.

Macken et al. (1999) emphasise the fact that different non-speech sounds can result in performance detriments on a serial recall task, as long as it contains “changing-state information” (Macken et al., 1999: 324), which allows for segmentation of a continuous signal into discrete chunks. This assumption is in line with the finding that serial recall is far more strongly disrupted by fluctuating than by constant broadband noise. Their *changing-state hypothesis*¹⁰ postulates that the primary task of serial recall in a typical ISE study (i.e. memorising and recalling verbal elements) competes for processing time on a *seriation mechanism*, which is concurrently engaged by obligatory and automatic auditory scene analysis processes that try to ‘parse’ fluctuating noise into auditory objects (Banbury, Macken, Tremblay & Jones, 2001; Jones, Madden & Miles, 1992; Macken et al., 1999). This kind of segmentation can be achieved by frequency (spectral) as well as intensity patterns. While it is difficult to compare the typical ISE task of recalling ordered lists of words to ‘normal’ speaking, it might be the case that keeping track of the serial order of elements in one’s own speech is a very basic, underlying function sub-serving the language production pro-

---

¹⁰The hypothesis has been framed within the object-oriented episodic record conception of working memory (O-OER; Jones, 1993). It differs from for instance Baddeley’s proposed working memory architecture (Baddeley, 2003a), by postulating a “global workspace”, on which abstract information objects or ‘chunks’ are linked through pointers. For an alternative account of the irrelevant speech effect, see for instance Neath (2000); however he claims to have no satisfactory principled explanation for the effect of non-speech noise.

---

9 Naturally, it would be a similarly interesting question whether an irrelevant sound effect affects processing-for-comprehension as well.
cess as well, including monitoring of previous speech. At first sight, this assumption might be at odds with observations made during speech intelligibility measurements, in which fluctuating noise is less detrimental than constant noise. However, this might have to do with the task in question: During processing-for-comprehension in a speech intelligibility measurement, listeners can rely on different cues to identify lexical material (and sentences). Insofar as serial ordering is a necessary component of sentence parsing as well, the presence of other cues might outweigh the lack or impairment of serial order information.

4.5. Summary

When comparing the linguistic abilities of different age groups if speakers with and without hearing impairment we have to deal with multi-dimensional interactions between factors like hearing status, age, environmental setting (silent or noisy), and the processing difficulty of stimulus material. The interactive effects of these factors might partially be mediated by cognitive or working memory ‘capacity’ constructs.

It is however still an open question, whether an adverse influence of noise on production is due to degradation of one’s own spoken input, or to noise perception itself, or whether both play a role and if so, how perceptual processing of noise and high-level processing of linguistic structure might interact. The dynamics of how external noise might interfere with the formulation process through the perceptual loop of the monitor or other mechanisms of interaction is currently understudied.

A possible construal of noise effects on language production is that listening to noise draws on resources otherwise required for language processing. If indeed linguistic processing at the level of formulation makes use of shared resources, at least to some extent, a noise-induced dual-task load might have a detrimental effect on processes involved in the generation of sentence structure. Following from the “interference-by-process” approach advocated by for instance Banbury et al. (2001), a detrimental effect on performance on the primary task should occur in cases where the (complex) behaviour required by the primary task draws on more basic functions also required by the secondary task of processing irrelevant sound (i.e. serialisation).
This hypothesis might create a valuable link to explain performance detriments when speaking in noise. The advantage of such an explanation is that it builds on highly explicit and psychologically plausible concepts for processing ‘capacity’. It also forms a puzzle piece for a broader picture of cognition, in which functionally specialised sub-processes interact in a particular way to process data of various kinds.

The studies reported in the following chapters will provide an empirical basis for answering the question which processing stages during sentence generation are susceptible to interference from something as unspecific as speech-free noise. As mentioned earlier in the introduction, this research will directly bear on the question of cognitive modularity in general, and modularity (and automaticity) of language processing in particular.
Part II.

Data
5. **Study 1: Near-spontaneous speech production**

5.1. **Introduction**

As mentioned in Chapter 4, different studies have researched the effect of noise on speaking before, mainly focusing on the acoustic effects on articulation. The study by Kemper et al. (2003) indicated that external noise might have an influence on other levels of language processing-for-production as well, in particular on the grammatical encoding stage (see Section 4.3.2 of Chapter 4). However, the compound measures for grammatical ‘complexity’ used by Kemper and colleagues have been considered rather coarse from a (psycho-)linguistic viewpoint (compare Hartsuiker & Barkhuysen, 2006). Therefore it might be of additional interest whether certain structures will be affected more than others by processing detriments related to background noise, or which aspects of sentence complexity are particularly difficult to process under adverse acoustic conditions. The current study aimed to explore the effects of noise on linguistic complexity of spoken language in greater detail, and in order to do so I searched for a number of different structure types as tokens of ‘difficult’ sentence structure.

5.2. **Experiment 1: Interview and picture story description**

5.2.1. **Spontaneous production**

For the current study, I elicited near-spontaneous speech samples from speakers under different conditions. Using spontaneous speech or conversation recordings has a couple of advantages over more strictly controlled methods for testing language production (cf. Eisenbeiss, 2010). First and
foremost the ecological validity of studies using naturalistic data is typically higher, because of the reduced artificiality of the task, compared to other elicitation methods. In addition, typically no (or much fewer) preconceived theoretical considerations can influence the design of stimuli. Of course, for the very same reasons, the conclusions that can be drawn from naturalistic or structured sampling are inherently limited, due to the lack of systematic experimental control (Bortz & Döring, 2006). Samples may vary greatly between subjects in size and content, and hence the linguistic contexts they might provide for particular phenomena can vary too (Eisenbeiss, 2010).

Despite these issues, naturalistic speech samples provide a good starting point for explorative research and hypothesis generation (Brown & Hanlon, 1970/2004). For the current study I compared speakers’ performance under different noise conditions within subject, and I did not perform comparisons between groups that systematically differ in properties inherent to the subjects. Therefore I should be able to interpret possible effects of noise causally, despite the limitations of the data collection method described before.

Based on the few earlier studies about sentence production in noise, I expected changes in voice intensity and/or quality (Lombard effect; see Section 4.3.2), speech and error rate, as well as a potential reduction in the amount of complex or difficult sentence structure. Because the current study was intended as one of the first steps to investigate the influence of a detrimental, noisy communication setting on syntax in language production, it is of exploratory nature. The research presented here might serve as a baseline for future studies testing other populations.

### 5.2.2. Measures for sentence structure complexity

Even to this day it poses a major problem for linguistic and psycholinguistic theory to establish generally accepted, ‘unified’, and empirically solid notions of complexity, see Chapters 2 and 3 for examples and discussion. For practical reasons, in the current study I have decided to search for different indicators of relative sentence complexity where I saw converging evidence from both psycholinguistic work and from theoretical assumptions about structural complexity based on formal linguistic analysis (see Section 3.4).
5.2. Experiment 1: Interview and picture story description

Embedding

The relative difficulty of embedded versus non-embedded clauses has been in the focus of much psycholinguistic research for decades, see for instance Lewis (1996) or Gibson (1998). Saffran et al. (1989) for instance use embedding as one indicator for structural complexity to assess the language production abilities of speakers with aphasia; Rosenberg and Abbeduto (1987) employ a similar criterion to assess production in children. The processing of structures with embedded clauses can be argued to be more difficult on semantic and syntactic grounds: The conceptual structure of such sentences is more complex, since the embedding relation must be encoded propositionally (e.g. Bloom et al., 1980; Schleppegrell, 1992; Yuasa & Sadock, 2002). On the syntactic level, formal relationships between matrix and embedded clause need to be represented, where different types of embedding (e.g. complement, adjunct, or relative clauses) might incur different processing difficulty, depending on the kind of relation that has to be processed (cf. Schütze & Gibson, 1999; Speer & Clifton, 1998; but also see Przepiórkowski, 1999), and depending on the depth of embedding (Delage & Tuller, 2010). For these reasons, I searched for tokens of embedded structure in the data.

Word order/canonicity

Other markers that have been argued to be indicative of structural complexity are the word order differences that let us distinguish for instance between SVO and OVS sentences in German or other V2 languages, between active and passive sentences, or between subject relative clauses and object relative clauses. Gorrell (2000) presented evidence for a subject-before-object preference in German, which is violated in sentences with a topicallyised object or in passives (but see Drai & Grodzinsky, 2006 for discussion of the status of passive sentences in German). Arguably, information structure and discourse context also play a role in how arguments are linearised and how much difficulty a particular linearisation poses during processing (Mak et al., 2008; Späth, 2003; Weskott, 2002). Nevertheless, without context the canonical word order can be considered ‘easier’ to process than a non-canonical word order. According to for instance Fanselow et al. (1999), what is particularly costly in ‘non-canonical’ sentences is the dislocation of
arguments and the concomitant difficulty in assigning the right theta-role to the right argument. Based on this premise I used the amount of instances of non-canonicity as one of the dependent measures.

**Other measures**

In addition to the structural indicators I counted the occurrence of grammatical errors and calculated measures for the Lombard reflex as well as for speech rate and fluency (measured in terms of *mean clauses per utterance* [MCU]), in order to replicate effects from earlier studies and to establish whether the background noise I used would have an observable effect on these aspects of language production.

**5.2.3. Method**

**Participants**

I tested 12 subjects, 7 female, between 20 and 30 years of age ($M = 23.4; SD = 3.15$). Subjects were recruited from among students of the University of Oldenburg and the Oldenburg Technical College, and they were paid 10 Euro per hour for participation. All participants reported to have no known history of language or speech disorders, or hearing impairments.

In order to estimate the role of differences between participants in terms of their cognitive capacities, I tested each subject with a *reading span* task (Daneman & Carpenter, 1980). This kind of test has been argued to be sensitive to individual differences in verbal working memory, especially with respect to tasks which put both storage as well as processing demands on language users (Just & Carpenter, 1992, but see Waters & Caplan, 1996). I tested subjects with German sentences modelled after the material used by Daneman and Carpenter (1980), and presented sentences in random order, as suggested by Friedman and Miyake (2005). A weighted score was obtained for each participant, taking into account the amount of correctly remembered sentence-final words per trial.

**Material**

Three kinds of stimulus material were used to elicit spoken language from participants: Two sets of interview questions, two short picture stories con-
5.2. Experiment 1: Interview and picture story description

containing six pictures, and long picture story containing 24 pictures. The interview questions revolved around the two topics holidays/festive days, and travel (see appendix A.1). The questions had been selected to be of a rather general nature in order to be applicable to a broad range of subjects. In addition, I selected two one-page picture stories from the Father and Son set of stories by German cartoonist e. o. plauen (Ohser, 2000). Both stories are made up of six black and white drawings without text, which depict individual scenes that connect in a self-contained plot. Finally, the long picture story I used was one of the text-free Frog Stories by Mercer Mayer (Frog, where are you?; Mayer, 1969), which I slightly shortened to fit 24 (2 × 12) single pages.

As distractor noise I used a speech-free sound signal that had been designed to model acoustic characteristics of six concurrent speakers (ICRA 7; Dreschler, Verschuure, Ludvigsen & Westermann, 2001). The signal possesses a frequency spectrum similar to that of speech. In addition, the intensity of the signal fluctuates (pseudo-)randomly over time, in order to imitate the 'babble' impression experienced in situations where many speakers talk at the same time (see 4). The noise signal is part of a standardised set of signals for audiological testing procedures, commissioned by the International Collegium of Rehabilitative Audiology.

Apparatus

Subjects were recorded in a sound-attenuated booth at the Speech and Music Lab of Oldenburg university. Recordings were made directly to hard disk, using a table-mounted AKG C-1000S microphone and an Echo Audio GINA 3G low latency sound adapter. The software used for recordings was PRAAT (Boersma & Weenink, 2011). The distance between a subject’s head and the microphone was approximately 60 cm. Participants sat in a chair which did not allow sliding or turning, but otherwise no restrictions were imposed on the speakers. Sound files containing the noise signal were played back through two Genelec 8020A loudspeakers, facing the subject at approximately 110 cm distance. In order to perform intensity measurements on the sound recordings, I calibrated the hard- and software setup using a Brüel & Kjær Investigator 2260 sound pressure level (SPL) meter. Root mean square (RMS) sound pressure level of the noise alone was 65 dB SPL at microphone position. Before every recording, the setup was checked for
correct position of microphone and loudspeakers.

Procedure

Subjects were tested individually. After briefing, a reading span score was determined for each subject. The recording session for the elicitation consisted of three parts, (i) a semi-standardised interview, (ii) two short picture story descriptions, and (iii) a long picture story description task. Each part was again divided into two halves, one half of the task carried out in silence, the other half in noise. The order of the two noise conditions was counterbalanced across subjects and tasks so that each half of every task was carried out under noise or in silence equally often. An entire session lasted about 45 to 60 minutes per subject, with an average net recording time of 15 minutes per subject (ranging between 10 and 32 minutes).

The interview was carried out by the experimenter, who sat opposite of the subject at a table inside the sound-attenuated booth while asking the interview questions. For the second task, each of the two short picture stories was presented on a single sheet of paper, and subjects were allowed to look at the entire story before starting to speak. The instruction given to the subjects asked them to narrate the story, rather than only describing the pictures’ content. The final task of the participants was to recount the wordless picture story Frog, where are you? (Mayer, 1969). In order to be able to record under the two noise conditions, the story was divided into two halves of 12 pages each. Every picture was presented on an individual page, and subjects turned the pages themselves. In this case subjects were not allowed to look ahead before turning the page. As with the short picture stories, the instruction was to narrate the events, rather than describing the picture contents.

Scoring and analysis

Sound recordings were orthographically transcribed by a specially trained transcriber, using the CHAT transcription standard (MacWhinney, 2000), and were verified by the author. Utterance boundaries were determined according to criteria given in the CHAT manual, based on prosodic, syntactic and semantic features, with priority of the former two types of features over the latter.
5.2. Experiment 1: Interview and picture story description

Since the recordings for each task differed in length between subjects, a sample was taken from each transcript that was approximately 300 (interviews), 150 (short picture story), or 400 (long picture story) words long for each half of the three tasks. Samples were drawn with the `kwa.1` command of the CLAN tools (MacWhinney, 2000) from the middle of each transcript and were extended to start and finish at utterance boundaries. For the acoustic analyses, the audio recordings were manually trimmed at the sample boundaries determined for the transcripts. For the measurements of intensity and speech rate I used the samples from our recordings, for which some additional signal processing and cleansing was necessary.

**Speech rate**  In order to obtain a speech rate measure for each sample file, I manually edited the audio files to remove experimenter speech from the interview recordings. I counted the amount of words spoken by the subjects in a sample using CLAN, and calculated speech rate in *words per second* (WPS) based on a measurement of the ‘cleaned’ sample audio file duration. For the speech rate measure, I kept all pauses made by the subject, since they obviously contribute to the speech rate construct.

**Intensity**  In order to test for an increase in vocal intensity, indicative of a Lombard effect, I calculated the *root mean square* (RMS) intensity of speech in both noise and silence. Since the calculation of a mean intensity value might be skewed by potential differences in the amount and length of pauses the subjects made in the two different conditions, I decided to remove all speech-free portions longer than 0.7 sec (in addition to all experimenter speech). The search for pauses was done with the help of a script for the PRAAT software by Lennes (2006). The results of the automatic search were manually verified, and the marked pauses were removed from the audio files. The RMS sound pressure in Pascal (Pa) was obtained for each processed sample audio file using PRAAT.

For the recordings I had opted for a free-field study setting, with noise exposure through loudspeakers. I considered this presentation method to be more natural than requiring subjects to wear headphones. For this reason, the sound recordings of the subjects’ speech also contain a noise signal portion. In order to estimate the ‘pure’ speech SPL in noise $L_s$, I measured the RMS sound pressure $p_n$ of a calibration recording that only contained noise.
and subtracted this value from the RMS sound pressure $p_{sn}$ I had measured for each recording of speech and noise combined. Accordingly, for recordings of speech in silence I subtracted an RMS sound pressure $p_n$ that was obtained from a silence recording with the same setup:

$$L_s = 10 \cdot \log \left( \frac{p_{sn}^2 - p_n^2}{p_{ref}^2} \right) \text{dB}$$

$p_s$, $p_{sn}$ and $p_{ref}$ were measured in Pascal (Pa); $p_{ref}$ is a reference value for calculating the sound pressure level of a sound event in decibels (dB), and is conventionally assumed to be $2 \cdot 10^{-5}$ Pa for sound travelling air.

**Complexity measures** Further analyses involving structure counts were carried out manually by the author, using the transcripts. A number of different indicators for structural complexity were obtained from the transcribed samples. Tokens of complex structure types according to the criteria given earlier, counted as indicators for complexity in production of subjects.

The value for *mean clauses per utterance* (MCU) takes into account the total number of main clauses and all embedded clauses in an utterance (e.g. Kemper et al., 1989; Nippold, Hesketh, Duthie & Mansfield, 2005), while a separate count of *embedded structures* contains the amount of dependent clauses only. In addition, I separately looked at three categories of structures from the embedded clauses: *relative clauses*, *complement clauses*, and *adverbial clauses* (cf. Delage & Tuller, 2010; Hamann et al., 2007).

The effects of noise on the canonicity of sentence structures in the produced language was assessed by counting two structures I believed to model cases of *non-canonical* structure: passive sentences and sentences with a topicalised (fronted) object. Finally, for the number of *ungrammatical structures* all clauses which contained grammatical errors (e.g. morphological errors, missing elements, or interrupted sentences) were counted.

**Design and statistical analysis** The presence of noise was counterbalanced across the different tasks and task halves, and was a within-subject factor. Results were analysed using mixed effects models (Baayen, 2008; Baayen, Davidson & Bates, 2008; Jaeger, 2008). To analyse the different
complexity measures, I calculated proportions (e. g. of embedded clauses relative to the total number of clauses), and computed linear mixed effects models with noise condition as predictor, using the \textit{lmer} package for the R statistical software. Random effects for subjects and task (interview, short picture story, long picture story) were estimated by adding random intercepts to the models. Intensity and speech rate measures were also evaluated with linear mixed effects models. P-values for the individual model parameters were calculated with the \textit{pvals.fnc()} function (Baayen et al., 2008).

In addition to estimating the effect of noise on subjects speaking, I entered their respective score on the \textit{reading span} task into the model, in order to check for correlations between each outcome variables and our measure for individual working memory capacity.\(^1\) As an additional step to analyse potential correlations with reading span, I performed a model comparison procedure as suggested by Baayen et al. (2008), in order to check whether the reading span measure would improve the fit of the respective model for each outcome variable.

\subsection*{5.2.4. Results}

Figures 5.1 through 5.3 show the results of the different measures in silence and noise respectively. The noise seems to have an effect on the vocal intensity with which the subjects spoke, and the amount of ungrammatical structures was significantly higher in noise than in silence. While there appear to be differences between the silence and the noise condition for other measures like MCU or the amount of non-canonical structures, these results were not statistically significant.

No correlation of subjects’ \textit{reading span} score with any of the different outcome variables reached significance. Table 5.1 provides an overview of the parameter estimates and p-values for reading span score as a coefficient in the different models. As the model comparisons for each outcome variable showed, adding the reading span score to the different models did not improve model fit in any case. Therefore, I removed this parameter from

\(^1\)While both parameters are estimated in one model, my interpretation of the two predictors used in the model differs in that I attempt causal inference from the noise factor, which was controlled and within-subject, whereas for the correlation between results on the reading span task and a particular result on an outcome variable I will not attempt a causal explanation in this study (Shadish, Cook & Campbell, 2002).
the model specifications. The parameter estimates for the effect of noise on the different outcome variables given in the following results are based on the simplified models.

**Table 5.1.** Correlations between weighted reading span measure and different outcome variables: coefficients for reading span as parameter in linear mixed effects models.

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Coef $\beta$</th>
<th>SE $\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>-0.406</td>
<td>0.88</td>
<td>-0.46</td>
<td>.782</td>
</tr>
<tr>
<td>Intensity (dB SPL)</td>
<td>4.833</td>
<td>3.96</td>
<td>1.22</td>
<td>.089</td>
</tr>
<tr>
<td>MCU</td>
<td>-0.136</td>
<td>0.29</td>
<td>-0.47</td>
<td>.615</td>
</tr>
<tr>
<td>Error count</td>
<td>-0.038</td>
<td>0.03</td>
<td>-1.39</td>
<td>.138</td>
</tr>
<tr>
<td>Non-canonical structures</td>
<td>-0.030</td>
<td>0.06</td>
<td>-0.52</td>
<td>.632</td>
</tr>
<tr>
<td>Embedded structures</td>
<td>-0.157</td>
<td>0.10</td>
<td>-1.53</td>
<td>.122</td>
</tr>
</tbody>
</table>

Panel A of figure 5.1 shows the average *speech rate* of subjects in noise and in silence. I do not find a statistically significant difference in speech rate between silence and noise ($N = 72$, log-likelihood = -28.47; Coef. = -0.002, SE = 0.07, $p = .980$). Panel B of figure 5.1 shows the result of the *intensity measurement* in silence and under noise. The effect of noise reaches significance: in noise, spoken language is about 8 dB (SPL) louder than in silence ($N = 72$, log-likelihood = -148.3; Coef. = 7.044, SE = 0.35, $p < .001$).

The average number of *clauses per utterance* (MCU) for the two noise conditions is given in panel C of figure 5.1. The MCU appears to be slightly higher in noise than in silence, but this difference does not reach significance ($N = 72$, log-likelihood = -2.728; Coef. = 0.098, SE = 0.05, $p = .071$).

Panel A of figure 5.2 shows the proportion of *ungrammatical clauses* out of all clauses produced by a subject for a given task, broken down by silence and noise. Subjects seem to produce slightly more ungrammatical structures under noise than in silence, which is indicated by a significant coefficient for the factor noise ($N = 72$, log-likelihood = 153.4; Coef. = 0.014, SE = 0.01, $t = 2.48$, $p < .05$). The grammaticality errors I observed manifested themselves in different ways. I categorised the different errors *post hoc* in order to check for noticeable patterns (see table 5.2 for an overview).

The proportion of *non-canonical structures* produced in silence and in noise is presented in panel C of figure 5.2. I do not find an indication that
5.2. Experiment 1: Interview and picture story description

Figure 5.1.: Speech rate (panel A), root mean square (RMS) intensity (panel B), and mean number of clauses per utterance (MCU; panel C), for silent and noise conditions.

Figure 5.2.: Proportion of ungrammatical clauses (panel A), embedded clauses (panel B), and non-canonical structures (panel C) in silence and noise.
Table 5.2.: Error counts in silence and noise, broken down by error type.

<table>
<thead>
<tr>
<th>error type</th>
<th>silence</th>
<th>noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexical errors</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>missing word</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>case errors</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>person, number, gender</td>
<td>—</td>
<td>4</td>
</tr>
<tr>
<td>word order</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>interleaved sentences</td>
<td>—</td>
<td>3</td>
</tr>
<tr>
<td>other</td>
<td>—</td>
<td>3</td>
</tr>
</tbody>
</table>

Noise has an effect on the amount of non-canonical structures ($N = 72$, log-likelihood = 86.35; Coef. = -0.018, $SE = 0.02$, $p = .275$). Panel B of figure 5.2 shows the proportion of embedded clauses produced in silence and under noise. The statistical model ($N = 72$, log-likelihood = 55.36) does not give reason to assume an effect of noise on the amount of embeddings (Coef. = 0.036, $SE = 0.02$, $p = .126$). For more fine-grained analyses, I broke down the total amount of embedded structures, to take a separate look at different kinds of clauses: relative clauses, complement clauses, and adverbial clauses (see figure 5.3). However, the statistical analyses did not yield significant differences between the silence and the noise condition for either of the three categories (all $p > .1$).

5.3. Discussion

5.3.1. Reading span

The results of this study yield no indication for a correlation between reading span and any of the complexity or fluency measures. This is in contrast to earlier studies, for authors such as Miyake, Carpenter and Just (1994) or Kemper and Sumner (2001) have reported such correlations and they have argued for reading span to be indicative of working memory capacity. WM capacity has been regarded by these authors as an important factor for creating and maintaining syntactic dependencies within sentences, one of the hallmarks of complex sentences. However, the validity of the reading span task as a measure for syntactic processing has been criticised by other au-
5.3. Discussion

Figure 5.3.: Proportion of embedded clauses by type: relative clauses (A), complement clauses (B), and adverbial clauses (C).

Authors, for instance Waters and Caplan (1996) who claim the reading span task is too different from processing for sentence comprehension (also see Friedman & Miyake, 2004). The lack of a significant correlation between reading span scores and the linguistic measurements taken in this study might support the latter point of view, although it should be borne in mind that my observations were on language production rather than comprehension. Given the divergent theoretical claims and empirical findings on this issue, I have to refrain from further speculation about why I did not find a correlation.

5.3.2. Lombard effect

The increase in speech intensity under noise appears to be a consequence of the noise condition created in the experimental setting. These results could be seen as indication that the current study successfully replicated the Lombard reflex or effect, in line with findings by for instance Lu and Cooke (2008) and a number of earlier studies on the effect.

In anticipation of concerns about the study setting, I have to issue a word of caution about the measurement method: Since I estimated the voice intensity of speakers from a sound recording which contained both the speaker’s voice as well as the distractor noise, the data might be subject
5. Study 1: Near-spontaneous speech production

to some measurement error.² But given the strong experimental control of hard- and software settings for the recordings, I am confident that the observed increase reflects an actual effect of noise. My trust in the data is supported by the fact that the magnitude of the effect I observed here (approximately 8 dB SPL) is similar to an earlier observation reported by Garnier, Bailly, Dohen, Welby and Lœvenbruck (2006), who found an increase of 8.6 dB SPL for utterances produced in (fluctuating) ‘cocktail party’ noise.

5.3.3. Speech rate

Different from what would have been expected based on earlier studies, I did not see an effect of noise on the rate at which subjects articulate. In the light of results by for instance Hanley and Steer (1949), Leder et al. (1987), Van Summers et al. (1988), Postma and Noordanus (1996), or my own results from the experiments reported in Chapters 6 and 7, I currently have no explanation for the lack of an effect.³ It is possible that the method I used for measuring speech rate comes with too broad a margin of error for obtaining a significant effect, either because of the necessary sample ‘cleaning’ steps or because a measurement in words per second is too coarse in scale.

5.3.4. Structural measures

The different measures of structural complexity do not indicate a systematic decrease of complexity under noise. This is in contrast to earlier findings by Kemper et al. (2003), and for now I can only speculate about the reasons for the difference in outcome.

²One anonymous reviewer and Martin Cooke (both reviewing an abridged version of this chapter for Language & Speech) have voiced concerns about the noise intensity level used in the study reported here. According to their comments, a voice intensity increase by about 8 dB SPL seems too high, given that distractor noise intensity was only 65 dB SPL. Martin Cooke further comments that in previous studies Lombard effects have been observed for noise intensities between 75 and 100 dB SPL. He suggests that the effect observed in my study might not be a Lombard effect, but might be attributable to the conversational setting, compare for instance Hazan and Baker (2011).

³The result is however compatible with the findings by Kemper et al. (2003), who observed changes in speech rate only with older subjects (see Section 4.3.2 of Chapter 4).
An obvious reason might be differences in terms of the experimental setting, for instance the type of noise that was used, and the intensity of noise presentation. However, the speech and noise distractor sounds used by Kemper and colleagues were presented at an intensity of 40-60 dB SPL. Despite the lack of further details on their experimental setting, this should amount to an effectively lower noise signal intensity than the intensity level used in the study reported here.

One other potential factor might be that the distractor noise used by Kemper et al. (2003) contained speech or speech fragments, very likely even in the “cafeteria noise” recording. In this case, the distractor stimulus would contain linguistic material that might arguably be more difficult to ignore than speech-free noise. Following the logic of dual-task experiments, if linguistic content is present, a competition might ensue between the processing of language for the (primary) speaking task and unconscious processing of speech in the distractor, resulting in capacity limitations on shared processes. Hence, I might speculate that in the setting used for this study there was too little ‘overlap’ between processing resources necessary to automatically process the noise signal and linguistic processing-for-production.

Additionally, the design of the study reported here might not have resulted in a conversational setting formal enough to warrant a speaking style that is typically characterised by a large amount of complex structure (cf. for instance Nippold et al., 2005). It is difficult to assess (and clearly beyond the scope of the present research project), however, in how far differences between studies in terms of lab setting or the wording of instructions might be able to affect the formality of the speech register.

In sum, I cannot find evidence for a ‘processing capacity’ effect of noise on speaking with respect to grammatical complexity, contrary to what was observed by Kemper et al. (2003). I can only speculate that the amount of distraction I offered to the participants was not strong enough, compared to the distractor stimuli used by Kemper and colleagues, even given the higher intensity level at which the noise signal was presented in the study reported here.
5. Study 1: Near-spontaneous speech production

5.3.5. Error count

The lack of conclusive results about sentence structure complexity notwithstanding, noise did affect the speech production of subjects in this study: The amount of grammaticality errors I observed was increased ever so slightly, but nonetheless significantly under noise. This is in line with earlier findings by Postma and Noordanus (1996), and complements their observations about phonological errors with errors on the morphological, lexical and syntactic level. Postma and Noordanus report a decrease in self-reported phonological error rates while producing tongue-twisters when overt, auditory feedback was suppressed for instance by (white, i.e. temporally unmodulated) noise. The noise signal was presented via headphones at 100 dB SPL, an intensity level which was certain to almost completely block external feed-back.

Postma and Noordanus (1996) attribute their finding of an increased error rate to the lack of an additional monitoring channel through the (external) auditory feedback loop. Based on Levelt’s perceptual loop theory, two mechanisms seem plausible to explain the error effects I observed, and both mechanisms are not mutually exclusive: First, some amount of acoustic masking of the signal that is fed back through the external loop will take place, and the monitoring of one’s own overt speech is less effective. Second, I might assume an effect of noise on the internal monitoring loop. While the language perception system is partially occupied with (automatic) processing of the distractor noise, the concurrent processing of internal loop information becomes impaired to the effect that erroneously specified parts of speech will be less likely to be intercepted before articulation. I will return to this point in more detail in the general discussion, see Section 8.3.3.
6. Study 2: Elicitation of complex sentence structure

6.1. Introduction

Study 1 yielded only limited evidence for an effect of noise on grammatical encoding, contrary to what had been expected from earlier work by for instance Kemper et al. (2003). A potential reason might have been the experimental task employed. Analysing (near-)naturalistic speech samples poses the risk of resulting in low numbers of tokens per structure category, and hence low statistical power. The second study therefore employed a more controlled experimental task and placed a narrower focus on processing of structures which have been argued to be ‘difficult’ (see Chapters 2 and 3). As exemplary cases of syntactic constructions which have been at the centre of much discussion with respect to complexity and processing difficulty, the work presented here will focus on two types of structure: (i) active and passive, or the grammatical voice diathesis (cf. Section 2.2.1), and (ii) relative clauses, in particular comparing clauses in which the relative pronoun acts as either subject or object (cf. Section 2.2.2). Of central interest to the study reported here was the influence of noise on the processing-for-production of these structures.

In addition, the study explored whether hearing status of the speaker, i.e. a possible hearing impairment, will influence the performance during sentence processing. Evidence from studies on language comprehension indicated that in some populations with impaired hearing, noise can have an adverse effect on processing, over and above mere perceptual masking effects. Studies by Wingfield and colleagues for instance concluded that processing of perceptually degraded language requires cognitive resources which are subject to inter-individual differences and age-related decline, and which are differentially taxed by the difficulty of the linguistic structure that has to be processed (McCoy et al., 2005; Wingfield et al., 2006,
6. Study 2: Elicitation of complex sentence structure

In particular, Wingfield and colleagues report multiplicative (or “superadditive”) effects between perceptual degradation, age and the structural ‘complexity’ of stimuli during perception experiments. The main goal of the study reported here was to assess the potentially interacting effects of age, noise and hearing impairment on the production of complex or difficult sentence structure.

The study thus tackled the question from two angles, using two different kinds of sentence structure and different populations. The goal was to find out whether background noise impacts the actual time course of sentence production, and to explore under which conditions older subject with hearing impairment show difficulty producing the more complex structures.

6.2. Experiment 2.1: Active and passive sentences

The first study tested the production of active and passive sentences. A number of different grammatical accounts for the passive rely on the intuitive assumption that the passive is in some way more difficult or more ‘marked’ than the active and try to capture this insight (with the possible exception of frequency explanations that are sometimes invoked in heavily lexicalised accounts such as Construction Grammar; see Chapter 2). What a great number of grammatical frameworks might agree upon is that analyses for the passive postulate more structure or additional computational steps to be taken in comparison to the ‘unmarked’ active case. *Ceteris paribus*, this additional structure should at least in principle lead to greater processing cost (cf e. g. Culicover & Jackendoff, 2005)—see Chapter 2 for discussion.

Since I did not expect that subjects with hearing impairment would generally have difficulties producing passive sentences at all, as sometimes observed in populations with language pathologies, I decided to find out whether production preferences for active sentences over the more complex passives changed depending on noise and hearing status. In order to elicit large enough numbers of passive sentences, I employed a picture description paradigm with written prime sentences, thus making use of structural priming effects in order to increase the likelihood of passive sentences being produced.
6.2.1. Structural priming

Structural priming refers to speakers’ tendency to sometimes ‘recycle’ sentence structure they have previously produced or perceived (Bock, 1986; Levelt & Kelter, 1982; Pickering & Ferreira, 2008). A number of studies since the early 1980ies has dealt with this phenomenon, see for instance Bock (1986); Bock, Dell, Chang and Onishi (2006); Chang, Bock and Goldberg (2003); Chang et al. (2000); Levelt and Kelter (1982), also see Pickering and Ferreira (2008) for a recent review.

Basically, structural priming studies try to control the conceptual message that subjects are supposed to produce in an experiment, for instance by having subjects describe pictures or by asking them to memorize and repeat a given sentence¹ after a delay. A sentence structure compatible with the message to be expressed is presented visually or auditorily as a prime structure. The dependent measure in most experiments is the relative frequency with which a particular structure is produced, contingent on the prime structure presented to the speaker; i.e. the relative number of cases in which subjects ‘recycled’ sentence structure.

Structural priming effects have been seen as evidence for an ‘autonomous’ syntactic processing level, or for the psychological reality of a syntactic representation. In order to explain the priming effect it has been suggested that structural priming facilitates the processing of a particular structure by prior activation of (lexicalised) pieces of syntactic structure representation (Pickering & Branigan, 1998), or through prior activation of mechanisms necessary for grammatical encoding (e.g. Chang et al., 2000). Hartsuiker and Kolk (1998) suggest, based on experimental findings from patients with Broca-type aphasia, that structural priming can actually help to alleviate difficulties in processing of complex or less frequent structure in special populations (see also V. S. Ferreira & Bock, 2006, for a review on the possible functions of structural priming).

For the elicitation study reported here, I pitted effects of priming, which should make production of passive sentences easier, against the assumed effects of structural complexity. Effects of noise on processing of complex

¹According to authors like Lombardi and Potter (1992); Potter and Lombardi (1990, 1998); Sachs (1967), speakers quickly forget surface structure in recall and memorisation tasks, and tend to not note changes in word order or syntactic structure in a recognition memory task after as little as 50 seconds Sachs (1967).
sentence structure then could manifest themselves in a change of the production rate for passives under the different priming conditions.

6.2.2. Method

Participants

I tested 24 native speakers of German, grouped by hearing status (with and without hearing impairment). The group of participants with hearing impairment consisted of 8 female and 4 male subjects, who were between 29 and 59 years old \((M = 45.7, SD = 11.3)\). In the group of participants with normal hearing were 3 male and 9 female subjects. They were between 30 and 48 years old \((M = 40.1, SD = 5.8)\). Subjects were recruited from a subject database of the Oldenburg “House of Hearing” audiological research centre and through public advertisements.

Subjects’ hearing level was determined with pure tone audiometry. Subjects in the group with hearing impairment were classified as having a slight/mild or moderate impairment (grade 1 or 2 according to WHO grading scheme; World Health Organization, n. d.), based on their pure tone average (PTA; averaging over hearing threshold at 500, 1000, 2000 and 4000 Hz) for the better ear. The PTAs in the group of subjects with hearing impairment ranged from 33.8 to 56.3 dB HL \((M = 45.2, SD = 8.0)\). Subjects who were wearing hearing aids were asked to remove them during the test session, because the different individual amplification settings are difficult to assess across the entire group. Normal hearing was assumed for subjects with a PTA of 25 dB HL or less in the better ear. Subjects in the group with normal hearing fell within a PTA range of between -2.5 and 10 dB HL \((M = 4.4, SD = 3.7)\).

In addition, I classified subjects with respect to their formal level of education, using a six-point scale based on the UNESCO International Standard Classification of Education (ISCED) scheme (United Nations Educational, Scientific and Cultural Organization, 1997/2006). A t-test comparing hearing impaired and normal hearing subjects did not reach significance, indicating that the average education level did not differ between both groups \((t(21.846) = -1.9612, p = 0.063)\).
6.2. Experiment 2.1: Active and passive sentences

Material

The stimuli consisted of 32 picture pairs, and matching prime sentences in German. Pictures and sentences were taken from the OLACS\(^2\) corpus of hearing test materials. This corpus contains a large number of sentences of varying syntactic complexity, which have been thoroughly controlled for lexical frequency and plausibility of the propositions expressed. A subset of 16 simple active (SVO) sentences from the corpus served as the basis for creating stimulus sentences and pictures. I chose sentences that expressed propositions which could be easily visualised, and which contained animate, anthropomorphic entities that would appear as highly recognisable characters in the picture. For each sentence I created the respective passive sentence version, including the subject of the corresponding active sentence in a prepositional phrase (see examples 16 and 17). In order to cue participants in the same way for active and passive sentences, as well as for filler sentences in other structures, I chose a VSO word order with a deictic adverb ("here") in topic position for all stimulus sentences.\(^3\)

(16) Hier interviewt der Tourist den Ritter.
    Here interview\textsubscript{3, sg} the\textsubscript{nom} tourist the\textsubscript{acc} knight
    "Here the tourist is interviewing the knight."

(17) Hier wird der Ritter vom Touristen interviewt.
    Here aux;\textsubscript{3, sg} the\textsubscript{nom} knight by-the\textsubscript{dat} tourist\textsubscript{dat} interview\textsubscript{ptcp}
    "Here the knight is being interviewed by the tourist."

\(^2\)Oldenburg Linguistically and Audiologically Controlled Sentences, cf. Uslar et al. (2013)

\(^3\)In the case of passive sentences, the verb in initial position is actually an auxiliary form, inflected for tense, person and number. The lexical verb is located at the end of the sentence and is realised as a participle in German. Arguably, these different ‘surface’ word orders could result in differences in processing between active sentences (lexical verb form bearing tense/inflection information at the beginning of the sentence) and passive sentences (tense/inflection information at the beginning, lexical content at the end of the sentence). In particular during sentence comprehension, processing differences are to be expected because the onset of processing grammatical and semantic information is different in both cases. See for instance De Goede (2007) for a comprehensive study on the time course of verb processing in Dutch. During production, a similar argument might be made based on the delayed onset of lexical verb retrieval. It is beyond the scope of the present study, however, to investigate the detailed implications of the differences in processing.
6. Study 2: Elicitation of complex sentence structure

Figure 6.1: Sample stimulus pictures from the OLACS corpus, showing simple transitive events (created by Albena Kaptebileva, see appendix B.1 for copyright information).

The stimulus pictures were black and white line drawings, showing a simple action between two animate entities (e.g. greeting, filming, or interviewing), and with a clearly recognisable assignment of thematic roles. Two versions of each picture were created, one showing the proposition of the stimulus sentence, and another one showing the same event, but with reversed thematic roles; compare panels a) and b) of figure 6.1 on page 116.⁴

A stimulus item consisted of a prime picture with a matching sentence describing the depicted event, and a target picture showing the same event and entities, but with reversed semantic roles. Each picture pair was combined with its respective prime sentence in passive and active, yielding 32 items in total (see appendix B.1 for a full list of stimulus sentences and pictures). I counterbalanced the horizontal orientation of the pictures, so that for half the items the agent would appear on the right. I also included 8 filler items in which both pictures were identical, so that subjects were not able to tell beforehand whether a target display would contain the exact same scene or a role-reversed variant. In total 40 prime-target pairs were combined with the stimulus material designed to elicit relative clauses for study 2, so in both studies stimulus items served as fillers for the respective other study.

⁴The propositions that the pictures were based upon had been tested for plausibility effects of role reversal with a questionnaire.
6.2. Experiment 2.1: Active and passive sentences

Figure 6.2.: Trial scheme for experiment 2.1: priming of active and passive voice.

Procedure

Testing was carried out in the communication acoustics simulator at the Oldenburg “House of hearing” audiological research centre. The simulator facility consists of a large sound-attenuated room with microphones and a multi-channel loudspeaker array hidden in the ceiling and in the walls. It is operated by a server system for sound processing in an adjoining control room.

After pre-testing and briefing, participants ran a short demonstration of the test procedure, in order to familiarise themselves with the task and its timing (see figure 6.2): An individual trial always started with a prime display, containing a prime image and a matching sentence in either active or passive voice. The prime was followed by the word “Hier...” (Here...) at the centre of the screen, which served as a cue for subjects’ responses. This was followed by a target display containing a picture only. Subjects had been instructed to inspect the first picture and read the sentence, and then describe the second picture in one sentence, using the sentence start cue that was given. Figure 6.2 gives the timing details for an individual trial.

Participants were tested individually, seated at the centre of the laboratory in front of a 17-inch LCD screen. An AKG C-1000 S microphone with cardioid recording characteristics was mounted on a table-top microphone.
stand and pointed towards the speaker. Participants’ answers were recorded to disk through a low-latency firewire audio interface (RME Fireface 400). Stimulus presentation and recording were controlled using the Experiment Builder software (SR Research Ltd., 2010). A software-implemented ‘voice key’ was used to measure the onset of a subject’s utterance and to determine the duration of recording. Maximum recording time was 11.5 seconds.

Two different acoustic conditions were created. A silent condition served as baseline, with no external noise in place. In order to create a background noise condition I presented speech-free ‘babble’ noise that has been designed to simulate the speech of six concurrent speakers in terms of (long-term) frequency spectrum and broad-scale intensity fluctuation. The ICRA 7 sound signal I used is part of a set of different speech-free noises created for the International Collegium of Rehabilitative Audiology (ICRA) for audiological testing procedures; see Dreschler et al. (2001) for details. Noise was presented through the lab’s built-in loudspeaker array, with settings calibrated so that the signal would reach an average intensity of 65 dB SPL at the approximate position of the subject’s head.

The two different acoustic background conditions I created during the experimental session were presented blockwise. In each noise block the full list of stimuli was used, in different pseudorandom orderings. Across participants, the order of the two acoustic settings was counterbalanced. Each of the two blocks (noise/no noise) took about 25 minutes to complete and included two pauses which the subjects could end on their own accord. The noise blocks were separated by a longer break which lasted at least three minutes. As a whole, a test session lasted between 60 and 80 minutes.

**Scoring**

The audio recordings of the subjects’ answers were transcribed orthographically and scored manually. Only fully recorded answers were analysed. An answer had to include the given sentence cue and two full noun phrases referring to the participants of the proposition in question. A sentence was scored as active if the agent was realised as a subject with nominative case marking and the patient or theme as an accusative object. Passive sentences

---

⁵The actual name participants used for the characters was not considered important, as long as the two different figures were unambiguously described.
6.2. Experiment 2.1: Active and passive sentences

had to contain the patient or theme in subject position, marked in nominative case and the agent in a prepositional phrase headed by an appropriate preposition (‘von’ or ‘durch’). Grammatical constructions that did not fulfil these requirements were scored as ‘other’; ungrammatical structures were scored as errors. In total, the amount of ungrammatical or incomplete structures produced during the priming task was fairly low, it amounted to only 11 cases out of 1536 critical trials. Because of concerns about statistical power no further analyses were carried out on the ungrammatical recordings.

Utterance latency was determined based on the voice-key measurements for the silence condition. Manual measurements of utterance onset with Praat (Boersma & Weenink, 2011) on a random sample of answer recordings confirmed that this method was reliable. For the noise condition, the onset latency was determined by manual measurements in all cases. Finally, in order to calculate speaking rate in syllables per second, I measured the total duration of each utterance with Praat, and divided it by the number of syllables in the actually produced sentence, which had been determined from the orthographic transcripts of the answer sentences.

Design and analysis

The study was based on a $2 \times 2 \times 2$ design, the factors being hearing status (2 levels, between-subject), prime structure (2 levels, within-subjects) and noise (2 levels, within-subjects). The data was analysed using a mixed effect modelling approach (cf. Baayen, 2008; Baayen et al., 2008; Jaeger, 2008). For count data (priming strength, count of passive sentences), logistic mixed effects models were established, for interval scaled data (utterance latency and speaking rate) I used linear mixed effects models. Utterance latencies were logarithmically transformed, in order to reduce skewness of the distribution. Fixed and random effect structure was evaluated with the anova() function in the R statistical software package (R Development Core Team, 2010). To check for possible correlations, age was added to the model as a continuous predictor. Significance values for the model parameter estimates given below were based on the outcome of the lmer() model fitting function for the logistic models. For the linear mixed models I used the pvals.fnc() sampling function (Baayen, 2008; Baayen et al., 2008).
6. Study 2: Elicitation of complex sentence structure

Figure 6.3.: Production of passive sentences modulated by hearing status, prime type, and noise.

6.2.3. Results

Priming strength

The results for the amount of passives produced as modulated by prime structure, noise and hearing status is shown in figure 6.3. Table 6.1 on page 121 summarizes the regression model used to evaluate the data. The significant negative intercept indicates that generally there is a clear preference for active over passive sentence structure, despite the priming manipulation. This general tendency interacts with prime voice: I found a strong increase in passive picture descriptions following a passive prime. As for the group differences, I saw that subjects with normal hearing appear to be only impacted by the manipulation of prime voice. In contrast, subjects with hearing impairment seem to behave differently under fluctuating noise, as evidenced by the significant three-way interaction between prime structure, hearing status and noise. The responses of subjects with hearing impairment show a larger variation between active and passive in silence. This large variation is reduced under noise to a level similar to that found among normal hearing subjects. Finally, age was only marginally correlated with prime voice. The older the participants, the more passives they produced in the passive priming condition.
6.2. Experiment 2.1: Active and passive sentences

Table 6.1.: GLMM model summary for passive priming strength (proportion of passives produced; $N = 1534$, log-likelihood = -580.3).

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.20</td>
<td>0.72</td>
<td>-5.9</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Prime voice:PASSIVE</td>
<td>5.15</td>
<td>0.93</td>
<td>5.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>1.07</td>
<td>0.69</td>
<td>1.6</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>Hearing status:HEARING IMPAIRED</td>
<td>2.03</td>
<td>0.82</td>
<td>2.5</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Age</td>
<td>-0.07</td>
<td>0.1</td>
<td>-0.7</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>Prime voice$\times$Noise</td>
<td>-0.35</td>
<td>0.75</td>
<td>-0.5</td>
<td>&gt;.6</td>
</tr>
<tr>
<td>Prime voice$\times$Hearing status</td>
<td>-2.42</td>
<td>1.14</td>
<td>-2.1</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Noise$\times$Hearing status</td>
<td>-2.85</td>
<td>0.84</td>
<td>-3.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prime voice$\times$Age</td>
<td>0.22</td>
<td>0.13</td>
<td>1.7</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Noise$\times$Age</td>
<td>0.08</td>
<td>0.09</td>
<td>0.9</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Hearing status$\times$Age</td>
<td>0.02</td>
<td>0.10</td>
<td>0.2</td>
<td>&gt;.8</td>
</tr>
<tr>
<td>Prime voice$\times$Noise$\times$Hearing status</td>
<td>2.47</td>
<td>0.93</td>
<td>2.7</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Prime voice$\times$Noise$\times$Age</td>
<td>-0.03</td>
<td>0.10</td>
<td>-0.3</td>
<td>&gt;.7</td>
</tr>
<tr>
<td>Prime voice$\times$Hearing status$\times$Age</td>
<td>-0.15</td>
<td>0.15</td>
<td>-1.1</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>Noise$\times$Hearing status$\times$Age</td>
<td>-0.06</td>
<td>0.10</td>
<td>-0.6</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>Prime voice$\times$Noise$\times$Hearing status$\times$Age</td>
<td>-0.06</td>
<td>0.11</td>
<td>-0.5</td>
<td>&gt;.6</td>
</tr>
</tbody>
</table>

Utterance latency

The time subjects needed to initiate their utterance was influenced by two factors: the sentence structure the subject was about to produce, and the presence of background noise (see figure 6.4a; table 6.2 on page 122 summarises the results of the statistical analysis). When planning an utterance in passive voice, it took on average 180 ms (about 12%) longer than with active voice utterances before articulation starts. With fluctuating background noise the initial latency was on average 80 ms smaller than in silence (about 5%). Finally, the order of noise condition blocks exerted a significant effect as well, utterance latency was reduced as subjects progressed further into the session (7.5% on average).

Speaking rate

For speaking rate, I saw a significant effect of the structure produced by participants: Passive sentences were articulated faster than active sentences by about 6%. See figure 6.4b and table 6.3 for details on this and the following results. Independently of this effect, hearing status exerted an influence
6. Study 2: Elicitation of complex sentence structure

(a) Utterance latencies modulated by structure and noise.

(b) Speaking rate modulated by structure and hearing status.

Figure 6.4: Utterance latency and speaking rate measures for active and passive priming.

Table 6.2: LME model summary for logarithm of utterance latency; N = 1464, log-likelihood = -146.6

<table>
<thead>
<tr>
<th>Coef β</th>
<th>SE (β)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.30</td>
<td>0.05</td>
<td>136.43</td>
</tr>
<tr>
<td>Produced structure:PASSIVE</td>
<td>0.11</td>
<td>0.02</td>
<td>6.80</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>-0.06</td>
<td>0.01</td>
<td>-4.23</td>
</tr>
<tr>
<td>Block order</td>
<td>-0.08</td>
<td>0.01</td>
<td>-8.25</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.41</td>
</tr>
</tbody>
</table>
6.2. Experiment 2.1: Active and passive sentences

Table 6.3: LME model summary for speaking rate in syllables per second (N = 1464, log-likelihood = -1464).

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE ($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.41</td>
<td>0.20</td>
<td>22.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Produced structure:PASSIVE</td>
<td>0.22</td>
<td>0.08</td>
<td>2.74</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>0.00</td>
<td>0.06</td>
<td>-0.01</td>
<td>&gt;.9</td>
</tr>
<tr>
<td>Hearing status:HEARING IMPAIRED</td>
<td>-0.57</td>
<td>0.26</td>
<td>-2.20</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Age</td>
<td>0.03</td>
<td>0.03</td>
<td>0.95</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Block order</td>
<td>0.11</td>
<td>0.02</td>
<td>4.60</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Produced structure×noise</td>
<td>-0.05</td>
<td>0.10</td>
<td>-0.46</td>
<td>&gt;.6</td>
</tr>
<tr>
<td>Produced structure×hearing status</td>
<td>-0.04</td>
<td>0.10</td>
<td>-0.35</td>
<td>&gt;.7</td>
</tr>
<tr>
<td>Noise×hearing status</td>
<td>-0.01</td>
<td>0.09</td>
<td>-0.15</td>
<td>&gt;.8</td>
</tr>
<tr>
<td>Produced structure×age</td>
<td>0.01</td>
<td>0.01</td>
<td>0.64</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>Noise×age</td>
<td>0.02</td>
<td>0.01</td>
<td>2.25</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Hearing status×age</td>
<td>-0.06</td>
<td>0.03</td>
<td>-1.80</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Produced structure×noise×hearing status</td>
<td>0.13</td>
<td>0.14</td>
<td>0.91</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Produced structure×noise×age</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.53</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>Produced structure×hearing status×age</td>
<td>0.00</td>
<td>0.01</td>
<td>0.33</td>
<td>&gt;.7</td>
</tr>
<tr>
<td>Noise×hearing status×age</td>
<td>-0.01</td>
<td>0.01</td>
<td>-1.40</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>Prod. structure×noise×hearing status×age</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.46</td>
<td>&gt;.6</td>
</tr>
</tbody>
</table>

on speaking rate, leading to significantly slower speech in the group of subjects with hearing impairment (14.5%). A simple effect for presentation block order shows that articulation became faster in the course of the session. Finally, there is a significant correlation between noise condition and age: the older a subject, the faster they articulated passive sentences under noise, as compared to production in silence.

6.2.4. Discussion

The results of experiment 2.1 demonstrate a clear effect of structural priming of the passive structure. The speakers’ tendency to re-use passive was markedly higher after a prime picture description in passive voice. This result is in line with previous demonstrations of active/passive voice priming, cf. for instance Bernolet, Hartsuiker and Pickering (2009); Bock (1986); Bock, Loebell and Morey (1992); Hanke (2007); Hartsuiker et al. (2004). The comparatively large magnitude of the priming effect for passive might be attributable to the fact that subjects could re-use lexical material for their
answer, including the verbs, in the trial procedure used. This ‘lexical boost’ has been shown earlier to influence priming strength (Melinger & Dobel, 2005; Pickering & Ferreira, 2008). Nevertheless, since thematic roles were consistently switched between prime and target pictures, the priming effect cannot be explained on the basis of lexical priming alone, but must also be based on a structural component.

The observation that utterance initiation times for passive sentences are significantly longer than for active sentences provides some indication that structural factors can influence the time course of utterance planning, see for instance F. Ferreira (1991) for similar evidence on other structures. However, since the passive sentences used here consistently contained an additional word compared to active sentences, it cannot be ruled out that the longer latency is due to an additional lexical retrieval process.

What is more, the observation that speaking rate is actually faster for passives than for active sentences at a first glance appears to speak against the assumption that greater syntactic complexity entails processing difficulty. However, I cannot exclude the possibility that this result is actually based on a systematic bias in the structures compared in this study. The passive sentences I elicited contained an additional auxiliary (‘wird’, the 3sg form of ‘werden’, to become), which tends to be articulated in a short and unstressed fashion, possibly resulting in on average faster articulation for passive sentences.

The reduction of spontaneous variation between active and passive picture descriptions observed under noise in the group of subjects with hearing impairment could be an indication that subjects with perceptual difficulty resort to active as the ‘simpler’ structural option under adverse circumstances. Given the comparatively small magnitude of the effect, this explanation remains somewhat speculative, however.

Subjects’ speaking rate also appears to be influenced by their age. The older a subject was, the faster he/she tended to articulate under noise. This tendency is different to what I would have expected from for instance Kemper et al. (2003), where the authors found older subjects to decrease fluency to cope with additional processing load. However, the task used here was markedly different in that the subjects were required to formulate and articulate a given message, with a primed sentence structure available. The increase in speed of articulation could reflect a (possibly strategically controlled) way to avoid forgetting of parts of the message. The marginally
significant interaction between age and prime voice suggests that subjects could rely more strongly on a potentially helpful information as they get older.

Across all other conditions, I see that hearing impairment leads to slower articulation, which is in line with the results of earlier studies: See for instance Goehl and Kaufman (1984); Postma and Noordanus (1996); Pratt (2004). This result suggests that hearing impaired subjects might resort to different strategies to cope with perceptual difficulties. I will return to this point in the general discussion in Chapter 8.

Finally, the effects of block order on utterance latency and speaking rate point to a training effect. Subjects became faster as the test session progressed. Because the order of the blocks of the different acoustic conditions was counterbalanced across participants, the noise effects reported above should be independent of noise condition block order.

To summarise, I did not find subjects with hearing impairment to have severe problems with producing the more difficult passive sentence structure. None the less, the results of this experiment suggest that subjects might rely on different strategies during on-line processing of complex structure, depending on the difficulty of the task and the individual resources available.

6.3. Experiment 2.2: Relative clause elicitation

The second experiment in study 2 aimed to elicit different types of embedded sentence structures, subject relative clauses and object relative clauses. Of immediate importance to the present study was to account for the difference in processing between subject and object relative clauses and how this difference might pertain to structural complexity differences (see Chapters 2 and 3).

Since (unlike passive sentence structure) relative clauses are difficult to prime, I employed a picture-based elicitation technique following Friedmann and Szterman (2006), to make subjects produce both subject and object relative clauses. With this method, participants did not have an actual ‘choice’ between two structures when describing the picture, as they arguably had with active and passive voice priming in experiment 2.1.
6. Study 2: Elicitation of complex sentence structure

6.3.1. Method

Experiments 2.1 and 2.2 were carried out in one combined session. The method was designed so that participants had to perform the same task for different types of sentences: first they saw a picture and read an accompanying sentence describing the picture, then they would get a cue for their answer sentence and see the target picture, which had to be described.

Participants

Participants were the same as in experiment 2.1.

Material

In total, 32 critical items were used, half of which were designed to elicit subject relative clauses, the others targeted object relative clauses. Every item consisted of a ‘model’ picture with a corresponding sentence describing the event, a sentence production cue, and a target picture which subjects had to describe.

As in experiment 2.1, the stimulus material was based on sentences from the OLACS corpus. 16 combinations of two propositions were chosen, which describe two concurrent events that can possibly happen at the same time: an intransitive action or state on behalf of one entity (to be described by a matrix clause), and a transitive action between this and two other entities (to be described by a subordinate relative clause). By varying the semantic role assignment for the transitive event between the two entities depicted, I could create two different sentences with embedded relative clauses in which the agent of the transitive (matrix clause) action would either bear the subject or the object role, marked by a nominative or accusative case form of the relative pronoun and respective agreement on the verb within the relative clause. For each of the 16 stimulus propositions, a corresponding subject and object relative clause version was created, compare examples 18 and 19.

(18) Hier zittert der Taucher, der die Zauberer malt.

Here tremble\textsubscript{3.sg} the\textsubscript{nom} diver relprn:nom the\textsubscript{acc.pl} magician\textsubscript{pl}
malt.

“Here the diver, who is painting the magicians, is trembling.”
6.3. Experiment 2.2: Relative clause elicitation

(a) Trembling diver is painting magicians
(b) Magicians are painting trembling diver

Figure 6.5.: Sample stimulus pictures for relative clause elicitation from the OLACS corpus (created by Albena Kaptebileva, see appendix B.2 for copyright information).

(19) Hier zittert der Taucher, den die Zauberer malen.

“Here the diver, whom the magicians are painting, is trembling.”

Based on the set of sentences, black and white drawings were created which showed the respective configuration of entities and event roles. Each picture was used two times during the session, once as a model picture and once as a target picture, each time in combination with the respective role-reversed picture. This resulted in 64 stimulus items in total (see appendix B.2 for a list of sentences and pictures). Across items, the horizontal orientation of the relative clause agents was counterbalanced. 16 filler items were created in which both pictures of the pair were identical, in order to minimise strategic behaviour of the subjects. The sentence production cues for the elicitation also contained the matrix clause verb, in order to restrict the answer options. The final test list containing 80 critical and filler items was used for both acoustic conditions, each time in a different pseudo-randomised order.

Procedure

The procedure was closely modelled after that of experiment 2.1 to give the impression of a uniform task across sentence types. Figure 6.6 gives an
6. Study 2: Elicitation of complex sentence structure

![Trial scheme for experiment 2.2: relative clause elicitation.](image)

Figure 6.6.: Trial scheme for experiment 2.2: relative clause elicitation.

overview over the timing of an individual trial.

**Scoring**

As with the answers from the first study, recorded picture descriptions were orthographically transcribed and manually categorised with respect to the sentence structure the subject had used. Only answers that had been completely recorded were evaluated. Ungrammatical sentences, as well as semantically incorrect renderings of the depicted scene, e.g. with reversed roles, were scored as errors. As in experiment 2.1, the amount of erroneous structures was comparatively low (32 sentences out of 1536 possible target sentences), therefore ungrammatical answers were not further analysed.

Target sentences had to include a matrix clause which started with the production cue (‘here’ matrix clause verb) and mentioned the agent of the intransitive action/event, and they had to contain a relative clause attached to the matrix clause subject. Relative clauses had to contain a relative pronoun, a noun phrase denoting the second entity and the verb. Depending on the case marking on the relative pronoun and on the second noun phrase within the relative clause, utterances were scored as containing either a subject or an object relative clause. During the scoring procedure, a third category was created *post hoc* for cases where subjects had produced an appropriate subject relative clause with passive, given that this structure is
available as an alternative to object relative clauses in German (see example 20). The occurrence of this kind of structure has been recently described by Belletti (2009) and Contemori and Belletti (in press) for Italian. Structures that included a different kind of embedding (e.g. causal or temporal adjuncts) or a coordination between two main clauses were scored as ‘other’.

(20) Der Hund, der von der Katze verfolgt wird, bellt.
The dog, who is being chased by the cat, is barking.

Finally, utterance latency and speaking rate in syllables per second were measured according to the same procedures as in experiment 2.1. For speaking rate, I calculated a value for each of the two clauses in all utterances containing a relative clause.

**Design and analysis**

Analysis of the interval scaled data from the utterance latency and speaking rate measurements proceeded as described for experiment 2.1. For the structure count data, two analyses were carried out, using logistic mixed effects modelling: First, I compared how inclined subjects were to avoid the intended target structure, with the count of structures produced as intended (subject or object relative clause) as dependent variable. The analysis design was similar to that of the priming study, with two within-subject factors (noise, with 2 levels, and intended answer structure, with 2 levels) and one between-subject factor (hearing status, 2 levels).

Because a high number of subject RC structures with passive were produced, I added a *post hoc* analysis, focussing on the more ‘difficult’ cases where the target utterance was intended to include an object relative clause. I compared how often the two major realisation possibilities, object RC and subject RC with passive, occurred as a function of noise conditions and hearing status (2 × 2 design).
6. Study 2: Elicitation of complex sentence structure

Table 6.4: GLMM summary of model for proportion of answers as intended (N = 1398; log-likelihood = -459.7).

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.75</td>
<td>0.62</td>
<td>2.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target structure: OBJECT RELATIVE CLAUSE</td>
<td>-5.24</td>
<td>0.45</td>
<td>-11.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Noise: ICRA7</td>
<td>0.48</td>
<td>0.43</td>
<td>1.1</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>Hearing status: HEARING IMPAIRED</td>
<td>-0.53</td>
<td>0.79</td>
<td>-0.7</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>Age</td>
<td>-0.03</td>
<td>0.04</td>
<td>-0.8</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>Block order</td>
<td>0.53</td>
<td>0.13</td>
<td>4.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Target structure x noise</td>
<td>0.06</td>
<td>0.56</td>
<td>0.1</td>
<td>&gt;.9</td>
</tr>
<tr>
<td>Target structure x hearing status</td>
<td>0.60</td>
<td>0.59</td>
<td>1.0</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Noise x hearing status</td>
<td>-0.54</td>
<td>0.55</td>
<td>-1.0</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Target structure x noise x hearing status</td>
<td>-0.35</td>
<td>0.75</td>
<td>-0.5</td>
<td>&gt;.6</td>
</tr>
</tbody>
</table>

6.3.2. Results

Avoiding object relative clauses?

Subjects more faithfully produced constructions which were intended to include a subject relative clause, as compared to targets that were intended to elicit object relative clauses. Apart from that, block order influenced the results. See table 6.4 for details on the statistics.

Object relative clauses and relative clauses with passive

In general, subjects produced more passive relative clauses than object relative clauses in those trials designed to elicit object relative clauses. Subjects’ choice between object relative clause realisations and realisations with passive subject relative clauses was also influenced by presentation block order, and crucially it was correlated with noise, hearing status and age. Figure 6.7 shows the correlation: Subjects with normal hearing appear to be more affected by noise than subjects with a hearing impairment, and with increasing age they tend to produce to a greater extent the (intended) object relative clause structure. This effect is greatly reduced for the hearing impaired subject group (see table 6.5 for a summary of the logistic regression model).
6.3. Experiment 2.2: Relative clause elicitation

Figure 6.7.: Interacting effects of hearing status, noise, and age on the substitution of object relative clauses with passives.

Table 6.5.: GLMM summary for amount of variation between object relative clauses and subject relative clauses with passive ($N = 601$, log-likelihood = -223).

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.97</td>
<td>0.73</td>
<td>4.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>-0.67</td>
<td>0.37</td>
<td>-1.8</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>Hearing status:HEARING IMPAIRED</td>
<td>0.65</td>
<td>0.88</td>
<td>0.7</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>Age</td>
<td>-0.24</td>
<td>0.11</td>
<td>-2.2</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Block order</td>
<td>-0.61</td>
<td>0.20</td>
<td>-3.0</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Noise$x$hearing status</td>
<td>0.53</td>
<td>0.54</td>
<td>1.0</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>Noise$\times$age</td>
<td>-0.12</td>
<td>0.09</td>
<td>-1.4</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>Hearing status$\times$age</td>
<td>0.15</td>
<td>0.12</td>
<td>1.2</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>Noise$\times$hearing status$\times$age</td>
<td>0.28</td>
<td>0.10</td>
<td>2.8</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
6. Study 2: Elicitation of complex sentence structure

Figure 6.8.: Correlation between hearing status, noise, age and utterance latency.

**Utterance latency**

As with the measurement of utterance latency in experiment 2.1, there was a significant effect of presentation block order (coeff. 0.09, SE = 0.04, t = 2.45, p < .05). In addition, a higher-order correlation between structure (subject relative clause with passive), noise, hearing status, and age reached significance (coeff. -0.02, SE = 0.01, t = -2.91, p < .01)—see table 6.6 on page 133.

**Speaking rate**

**Matrix clause** Speaking rate for the matrix clauses slightly increased with the progression of the recording session. Crucially, participants’ hearing status had an effect on the speed of articulation for the matrix clause portion, if an object relative clause was to follow. In this case hearing impaired subjects showed a slower speaking rate than normal hearing subjects (see figure 6.9; for a summary of the results of the statistical analysis, see table 6.7 on page 135). Both noise condition as well as the structure about to be produced in the relative clause appear to be correlated with age in
Table 6.6: Summary of LME model for log utterance latency ($N = 1161$, log-likelihood $= -131.6$).

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef $\beta$</th>
<th>SE ($\beta$)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.46</td>
<td>0.09</td>
<td>86.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>produced structure:OBJECT RELATIVE CLAUSE</td>
<td>0.04</td>
<td>0.06</td>
<td>0.68</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>produced structure:PASSIVE RELATIVE CLAUSE</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.31</td>
<td>&gt;.7</td>
</tr>
<tr>
<td>noise:icra7</td>
<td>-0.12</td>
<td>0.03</td>
<td>-3.98</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>hearing status:HEARING IMPAIRED</td>
<td>-0.12</td>
<td>0.11</td>
<td>-1.05</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>age</td>
<td>0.02</td>
<td>0.01</td>
<td>1.56</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>block order</td>
<td>-0.10</td>
<td>0.01</td>
<td>-9.48</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×noise</td>
<td>0.09</td>
<td>0.07</td>
<td>1.29</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure:PASS. RC×noise</td>
<td>0.06</td>
<td>0.05</td>
<td>1.16</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×hearing status</td>
<td>-0.01</td>
<td>0.08</td>
<td>-0.10</td>
<td>&gt;.9</td>
</tr>
<tr>
<td>produced structure:PASS. RC×hearing status</td>
<td>0.08</td>
<td>0.05</td>
<td>1.79</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>noise×hearing status</td>
<td>0.12</td>
<td>0.04</td>
<td>2.88</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×age</td>
<td>-0.02</td>
<td>0.01</td>
<td>-2.07</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>produced structure:PASS. RC×age</td>
<td>-0.01</td>
<td>0.01</td>
<td>-2.78</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>noise×age</td>
<td>-0.01</td>
<td>0.00</td>
<td>-2.22</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>hearing status×age</td>
<td>-0.02</td>
<td>0.01</td>
<td>-1.32</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×noise×hearing status</td>
<td>-0.17</td>
<td>0.11</td>
<td>-1.56</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure:PASS. RC×noise×hearing status</td>
<td>-0.07</td>
<td>0.07</td>
<td>-1.10</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×noise×age</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.69</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>produced structure:PASS. RC×noise×age</td>
<td>0.02</td>
<td>0.01</td>
<td>2.38</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×hearing status×age</td>
<td>0.01</td>
<td>0.02</td>
<td>0.37</td>
<td>&gt;.7</td>
</tr>
<tr>
<td>produced structure:PASS. RC×hearing status×age</td>
<td>0.02</td>
<td>0.01</td>
<td>2.97</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>noise×hearing status×age</td>
<td>0.02</td>
<td>0.01</td>
<td>3.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>produced structure:OBJ. RC×noise×hearing status×age</td>
<td>0.04</td>
<td>0.02</td>
<td>1.72</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>produced structure:PASS. RC×noise×hearing status×age</td>
<td>-0.02</td>
<td>0.01</td>
<td>-2.92</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
the subject group. With increasing age, speakers tended to articulate passive relative clauses more slowly than subject relative clauses. Under noise, generally, the opposite correlation appears to hold: the older a speaker, the higher the speaking rate in noise.

**Relative clause** The speaking rate on the embedded relative clauses was different only for RC structures with passive, which were articulated faster (coeff. 0.38, $SE = 0.16$, $t = 2.33$, $p <.05$).

### 6.3.3. Discussion

A methodological concern has to be raised about an effect of experimental design: as in the first experiment, some effects of block order reached significance, which are evidence for a training effect. Given the counterbalanced design, however, these effects should be independent of other effects concerning noise, sentence structure or individual differences.

Generally, subjects showed a tendency to avoid object relative clauses. This fact provides evidence for the common assumption that this particular structure is more difficult to produce than subject relative clauses. Appar-
Table 6.7: Summary of LME model for matrix clause speaking rate in syllables per second ($N = 1161$, log-likelihood = -1499).

<table>
<thead>
<tr>
<th></th>
<th>Coef $\beta$</th>
<th>SE ($\beta$)</th>
<th>$t$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.29</td>
<td>0.29</td>
<td>15.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>-0.59</td>
<td>0.23</td>
<td>-2.51</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>-0.08</td>
<td>0.19</td>
<td>-0.44</td>
<td>&gt;.6</td>
</tr>
<tr>
<td>noise: icra7</td>
<td>0.12</td>
<td>0.10</td>
<td>1.15</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>hearing status: hearing impaired</td>
<td>-0.62</td>
<td>0.35</td>
<td>-1.77</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>age</td>
<td>0.04</td>
<td>0.04</td>
<td>0.96</td>
<td>&gt;.3</td>
</tr>
<tr>
<td>block order</td>
<td>0.09</td>
<td>0.04</td>
<td>2.44</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>0.35</td>
<td>0.23</td>
<td>1.50</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>0.25</td>
<td>0.17</td>
<td>1.52</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>0.58</td>
<td>0.28</td>
<td>2.12</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>0.29</td>
<td>0.15</td>
<td>1.87</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>noise x hearing status</td>
<td>-0.10</td>
<td>0.14</td>
<td>-0.70</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>-0.05</td>
<td>0.03</td>
<td>-1.48</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>-0.04</td>
<td>0.02</td>
<td>-2.08</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>noise x age</td>
<td>0.03</td>
<td>0.02</td>
<td>2.15</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>hearing status x age</td>
<td>-0.06</td>
<td>0.05</td>
<td>-1.46</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>-0.63</td>
<td>0.36</td>
<td>-1.74</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>-0.43</td>
<td>0.22</td>
<td>-1.95</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>0.10</td>
<td>0.06</td>
<td>1.70</td>
<td>&lt;.1</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>0.03</td>
<td>0.02</td>
<td>1.06</td>
<td>&gt;.2</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>0.03</td>
<td>0.05</td>
<td>0.59</td>
<td>&gt;.5</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>0.03</td>
<td>0.02</td>
<td>1.52</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>noise x hearing status x age</td>
<td>-0.03</td>
<td>0.02</td>
<td>-1.47</td>
<td>&gt;.1</td>
</tr>
<tr>
<td>produced structure: obj. rel.</td>
<td>-0.05</td>
<td>0.07</td>
<td>-0.75</td>
<td>&gt;.4</td>
</tr>
<tr>
<td>produced structure: pas. rel.</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.37</td>
<td>&gt;.7</td>
</tr>
</tbody>
</table>
ently it also tends to be more difficult than subject relative clauses with passive, as the preference for the latter structure over object-initial relative clauses suggests. This is in line with previous findings by for instance Belletti (2009); Contemori and Belletti (in press). Certainly, it has to be taken into account that for subject relative clauses no actual alternative option is available, which is arguably the case for the variation between object relative clauses and passive subject relative clauses. Still, the higher number of descriptions produced with a passive subject relative clause structure instead of an object relative clause structure is a clear indicator of speakers’ preference for the more ‘canonical’ order of grammatical roles. A reason might be that the relative clause construction with passive allows to maintain the topic status of the matrix clause element that is being modified and no intervening or interfering information has to be assigned to elements (Belletti & Chesi, 2011). However, it still remains to be explained how such a structural difference results in a preference for one type of structure over the other in a given situation.

A different aspect is that priming from earlier trials either containing a passive prime sentence or requiring the subject to produce a passive sentence might have led to the high amount of passive relative clauses, by making the passive more available to speakers. However, the substitution of the (intended) ORC with a passive SRC occurred spontaneously—during the study the passive relative clauses structure never appeared as a model sentence or in the instructions. What is more, there were no cases in which subjects opted to replace a subject relative clause with an object relative clause with passive. This is a viable structural choice and should be expected to occur if the ‘contamination’ of relative clause structures with passive were due to priming. Hence I assume that subjects’ preference for the passive subject relative clause structure is linked to ease of processing of sentence structure. What is more, the trials designed to elicit subject relative clauses contained a model sentence with a real object relative clause, so arguably I should have been able to observe priming of the OVS word order in those embedded sentences as well.

Despite what I would have expected from the data by F. Ferreira (1991), there was only a small effect of structural complexity on utterance latency, and this effect manifested itself in a higher-order correlation. The data suggest that with increasing age subjects from the hearing impaired group became faster when producing passive subject relative clauses under noise.
At present, no straightforward explanation for this interaction seems available. It should be borne in mind, however, that the task used here differed from that employed by F. Ferreira (1991): I asked subjects to prepare an utterance under time pressure, crucially starting with visual information rather than memorising pre-formulated sentences. Since language production is usually assumed to proceed in an incremental fashion (Levelt, 1989), the structural differences between subject and object relative clauses (and subject relative clauses with passive respectively) might not come into play to influence the timing of processing until after articulation of the matrix clause has started. After all, the matrix clause structure was identical for all conditions.

And as the results of the speaking rate analysis suggest, at least in the group of subjects with hearing impairment the speed of articulation on the matrix clause is indeed influenced by the structure that is about to follow. The slower speaking rate on the clause before an object relative clause or a passive subject relative clause should reflect differences in the planning processes that can be linked to the more complex structure. I did not find a reduction in speaking rate on the object relative clause portion itself. To me this suggests that the more ‘demanding’ structural planning processes for the respective relative clause might already be over at this point in time.

Similar to the result from experiment 2.1, there was an increase in speaking rate for the passive sentence in cases where a subject relative clause with passive had been produced. As with the finding about the speaking rate for passive sentences from experiment 2.1, I cannot exclude the possibility that the additional function word in passive sentences, which is usually pronounced in a fast, unstressed fashion, might have created a bias towards faster speaking rate measurements.

Furthermore, the observation that hearing impaired subjects show a reduced speaking rate on matrix clauses before object relative clauses and subject relative clauses with passive might reflect a compensation strategy. Since the realisation of a second proposition in a complex sentence structure might be particularly taxing, the decrease of articulation speed could help to prevent forgetting of relative clause content. Also possibly attributable to strategic differences is the interactive effect of hearing status, noise and age. Subjects with hearing impairment might rely on the ‘simpler’ passive subject relative clause structure across all age ranges, whereas subjects with normal hearing tend to stick with the model object relative clause.
construction more often as they get older. However, this interpretation is based on a correlation with the covariate age, and therefore has to be taken with a grain of salt. The correlation results further suggest that with increasing age, speaking rate on matrix clause before the more complex passive subject relative clause decreases. This result would be in line with findings by Kemper et al. (2003), but stands in contrast with the findings from experiment 2.1. At present, I have no explanation for this difference.

In summary, and similar to experiment 2.1, the effects observed here do not yield straightforwardly interpretable interactions, such as those that have been shown in studies testing language comprehension measures. I did find some indication that individual differences with respect to hearing status and age have an influence on the language production process for complex structure under adverse conditions. I will return to these results in the general discussion in Chapter 8.
7. Study 3: Production of number agreement¹

7.1. Introduction

The previous two studies researched the effects of noise on the production of ‘difficult’ syntactic structure at sentence level. The effects reported present a (not always clear) picture about the types of structure affected and rough magnitudes of the influence on the speaking process. However, the two studies still fall short of examining more closely a mechanism for the interaction between the processing-for-production of sentence structure and listening to noise. Therefore, I opted to take a closer look at the production of agreement between subject and verb as a testing case, and examined the influence of external noise on the amount of agreement errors. Agreement was of particular interest, because it should be considered part of the grammatical encoding stage, during which grammatical functions are assigned and constituent structure is assembled. This processing stage is typically assumed to proceed in a largely automatic, ‘modular’ fashion (see Chapter 3).

In order to test the influence of situational, external distraction on agreement generation, I conducted a study employing an adapted elicitation paradigm in which subjects had to produce sentences under noise, based on sentence preambles they were given as stimuli. In order to avoid noise masking during the perception of the stimuli, I opted for visual presentation of the preambles. Based on the literature reviewed in Chapter 4, I expected that speech-free noise will create a secondary task load that can interfere with language processing. The findings from the ISE literature

---

¹An extended version of this chapter has been published as a journal article: Hanke, M., Hamann, C., & Ruigendijk, E. (2013). On the laws of attraction at cocktail parties: Babble noise influences the production of number agreement. *Language and Cognitive Processes*, 28(8), 1114–1133. (Copyright Taylor & Francis)
further predicted that temporally structured (fluctuating) noise would lead to stronger disruptions than constant noise. I therefore used four different noise conditions: a silent baseline, steady-state noise, and two fluctuating noise signals which differed in terms of the intensity of the fluctuation.

I am at present unaware of any previous studies that investigated the effects of external noise as a distractor on the production of agreement with the paradigm employed here. Therefore I had no reference point for a quantitative hypothesis about the effect of noise. Still, I expected the additional load generated by the presence of noise to impact the language production process. It was assumed that secondary task load effects are the result of a competition for processing time on procedures shared between primary and secondary task. Therefore I expected a slowing down of language processing, which in turn should be reflected in higher error rates. In addition, on-line measures like utterance latency or speech rate should be affected, leading to longer latencies before speaking starts, and to slower articulation.

7.2. Pre-test: Questionnaire study

7.2.1. Method

In order to verify the material designed for this study and to check for plausibility effects, I carried out a pre-test with a written sentence completion task. The material was pre-tested in an online questionnaire that allowed to test a large number of subjects in a short time span.

Participants

The questionnaire was completed by 90 subjects, aged between 19 and 59 years ($M = 24.6$, $SD = 5.21$). 23 participants were male, 67 were female, and all reported themselves to be native speakers of German. Subjects were recruited from the student and staff body of Oldenburg university through an on-line bulletin board announcement. They participated voluntarily, without receiving compensation. In addition to the 90 subjects who completed the form, 122 persons started the questionnaire but did not finish, or declared they were no native speakers of German.
7.2. Pre-test: Questionnaire study

**Material**

27 German test sentences were constructed after the pattern used by Bock and Miller (1991); they consisted of a copula construction with an embedded prepositional phrase (PP) modifying the subject/head NP (see also Hartsuiker & Barkhuysen, 2006; Hemforth & Konieczny, 2003; Schriefers & van Kampen, 1993). In cases where appropriate, German translations of material from earlier studies like Hartsuiker and Barkhuysen (2006) were used. The (intended) structure of the sentences was \[ S[NP D N[PP P \[NP D ADJ N]]] V AUX ADJ \], see example 21. All head nouns used in the material had feminine gender. This was to avoid any additional cue from the determiner, which is usually marked for case, gender and number in German. The paradigm for the definite determiner shows syncretism between singular and plural for feminine gender in nominative and accusative case, hence the determiner would invariably have the form “die”. For the embedded attractor phrase I used prepositions which select for either accusative or dative case, and I combined them with nouns of masculine, feminine or neutral gender. An adjective was added to the intervening prepositional phrase, in order to increase the distance between subject noun and verb. With the chosen combinations of preposition and noun, the determiner of the embedded noun was always unambiguously marked for singular or plural,\(^2\) and therefore the form of the adjective did not change, which would be the case if combined with an ambiguously marked definite determiner. Additionally, some of the resulting noun-combinations allow for a distributive reading, which has been shown to increase error rates (Vigliocco et al., 1996).\(^3\) See appendix C for a list of all experimental sentences.

\(^2\)Incidentally, the determiners were in most cases also unambiguously marked for (non-nominative) case. As one of the anonymous reviewers for Hanke, Hamann and Ruijghendijk (2013) rightly pointed out, these linguistic factors can influence the likelihood of agreement errors (Hartsuiker et al., 2003). The main focus of the present experiment lay on the influence of noise on agreement errors, and noise was a within-item factor in our design. Therefore I will leave the question after potential *interactions* between the effects of noise and different linguistic cues for agreement computation open for future research.

\(^3\)As with case marking on the determiner, the semantic or conceptual difference between sentences with and without distributive reading was not central to the research goal of this paper. All sentence preambles were used under all noise conditions, and I will leave the question about potential interactions between noise and distributivity open at present.
For each of the sentences, two nouns and respective adjectives were combined to form plausible propositions. Nouns and adjectives used in the sentences were controlled for word length in syllables, as well as for frequency class according to the Leipziger Wortschatz database (Biemann, Bordag, Heyer, Quasthoff & Wolff, 2004), so that lexical elements within an individual sentences would not stand out on either of the two measures. The number marking on the test sentence NPs was manipulated to yield four different versions per sentence. In order to do so, I combined the two factors match (subject and local noun match/mismatch in number) and subject noun number (singular/plural).

The resulting 108 test items were distributed over four lists, so that each list contained 27 unique combinations of two nouns. Subject number/match conditions were equally distributed across lists. 42 filler sentences were added, taken from the OLACS corpus of sentence stimuli (Uslar et al., 2013). They contained simple transitive sentences (see example 23) and sentences with embedded relative clauses (see example 24). Critical and filler sentences were pseudo-randomised, each list starting with at least six filler sentences and with no more than two critical sentences following each other.

(23) Der freche Kasper tadelt den stolzen Clown.
    "The cheeky buffoon is reprimanding the proud clown."

(24) Der Soldat, der die Köchinnen tadelt,
    sweats.
    "The soldier, the chefs are reprimanding,"
7.2. Pre-test: Questionnaire study

“The soldier, who is reprimanding the female chefs, is sweating.”

In critical sentences, the gap for the sentence completion would always take the position of the auxiliary verb, the penultimate word of the sentence. In filler sentences, the position of the gap was varied to allow for full verb or adjective fillers in different positions towards the end of the sentence. This was to prevent any linear position or word class from standing out prominently across stimulus lists.

Finally, an additional four stimulus lists for the plausibility rating were created. Each of the four lists contained all 108 test sentences in different randomisations.

Procedure

The study was carried out as an online questionnaire, using the LimeSurvey web application (Schmitz, 2009). First, subjects had to carry out the sentence completion task. Stimulus sentences were presented one at a time, on a single line with underscores indicating the gap position. Placed underneath it was a text input field for the participant’s answer. After providing a completion, participants could move on to the next item. Subjects were instructed to fill in a single word that would complete the sentence in a grammatical fashion. The instructions told subjects to work through the questionnaire in one go without pauses, however there were no restrictions on how long subjects took to fill out the entire set of questions. Participants were also instructed not to think too long about their answers and not to consult with others.

After finishing the sentence completion task, participants were presented with items for the plausibility rating. All 108 critical sentences were presented, this time in a completed form, including a form of the auxiliary verb to be. Participants were asked to read each sentence and rate it on a four-way scale ranging from not plausible at all to very plausible. The instructions further stated that all sentences could be considered grammatical and that subjects should base their ratings on content; the term plausible was paraphrased as “traceable, reasonable, comprehensible, and perspicuous”. Again, each sentence was presented on a single page, with a radio button selection field for the rating. Subjects had to make a decision on a sentence before they could proceed.
After the rating block a few demographic questions remained before the questionnaire ended. It took participants approximately 25 minutes to complete the entire survey.

Scoring

Only questionnaires that had been fully completed were subjected to further analysis. The answers given in the cloze task were scored manually: An answer was scored as an agreement error if the verb’s number did not match the number of the subject noun. Since participants had to provide an answer before they could continue, there were no omissions. Only in very rare cases the word given as an answer was not a verb or would not form a valid completion of the sentence; these cases were removed from the analysis. The scores for the plausibility rating were calculated by averaging values assigned internally to the rating scale, with *not plausible at all* being equivalent to a score of 1 and *very plausible* being equivalent to a score of 4.

Design and Analysis

Plausibility rating results were used to eliminate those stimuli from the set where the variation between the four different number combinations was too high. I removed five items where the mean plausibility ratings differed significantly (t-test) between the four number combinations, and two additional items where the standard deviation across all ratings was higher than 0.9. The error rates on the remaining 80 sentences/20 items were analysed using a generalised logistic mixed model approach (GLMM; cf. Baayen, 2008; Jaeger, 2008), with subjects and noun-noun combinations as random effects, and match and subject noun number as (within-subjects) fixed effects.

7.2.2. Results

Table 7.1 shows the coefficients of the GLMM model. There was a simple effect for number marking on the subject, indicating that the amount of errors is larger when the subject was marked plural. I also found a simple effect of *number mismatch*: when subject and local noun differ in number
Table 7.1: GLMM coefficients for the item pre-test error data (predictors are treatment-coded).

<table>
<thead>
<tr>
<th></th>
<th>Coef</th>
<th>SE(β)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.67</td>
<td>0.49</td>
<td>-7.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Match</td>
<td>1.31</td>
<td>0.64</td>
<td>2.0</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Number</td>
<td>1.78</td>
<td>0.64</td>
<td>2.8</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Match × Number</td>
<td>-1.33</td>
<td>0.87</td>
<td>-1.5</td>
<td>&gt;0.1</td>
</tr>
</tbody>
</table>

marking, the amount of errors was higher (see figure 7.1). The interaction effect between the factors number and match did not reach significance.

Figure 7.1: Error percentage by subject number and match condition.

7.3. Experiment 3: Agreement error elicitation

7.3.1. Method

For the experiment I used a sentence elicitation task with *rapid serial visual presentation* (RSVP) of sentence preambles (Potter, 1984). The visual presentation was necessary because part of the presentation would take place under noise, and the noise would partially mask auditorily presented stimuli,
introducing an additional, severe source of error to the subject’s answers. According to Potter (1984), the serial word-by-word presentation mode is more similar to listening than to conventional reading, given that looking ahead or back is not possible with this way of presentation. Results from studies, for instance by Potter and Lombardi (1998), indicate that subjects can accurately read and recall sentences shown with RSVP.

Participants
24 undergraduate and master students of the University of Oldenburg, aged between 20 and 28 years ($M = 23.6$, $SD = 2.12$), participated in the study; half of them were female. Subjects were recruited through an on-line bulletin board announcement, and were paid 7.50 Euro per hour for participation. All participants were naive with regard to the experimental manipulation. Before testing, each subject was screened for hearing acuity with pure-tone audiometry, carried out on an Interacoustics AC40 audiometer. All participants had normal hearing according to WHO (World Health Organization, n. d.) standard. In order to increase statistical power, each participant was tested two times, with at least two weeks in between sessions. Only data from participants that had taken part in both experimental sessions was used for the analysis.

Material
20 test sentences were selected from the materials that had been tested in the questionnaire study. Based on the rating results from the pre-test, I removed items where the difference in plausibility between the four different head/local NP number combinations reached significance, and items where the overall standard deviation was above 0.9. This left 20 experimental items, with four sentence versions per item (see appendix C for a list of items). For the experiment, the four number marking conditions (singular match [sg-sg], plural match [pl-pl], singular mismatch [sg-pl], plural mismatch [pl-sg]) were counterbalanced across four noise condition blocks, so that one combination of nouns never appeared more than once within one block. Four experimental lists were created with different distributions of material across blocks.

In addition, the experimental lists were pseudo-randomly interspersed
with filler items taken from the OLACS corpus of stimulus sentences (Uslar et al., 2013). Care was taken to limit the number of succeeding experimental items to three in all cases. The filler material consisted of 160 sentences, 80 relative clauses and 80 simple transitive clauses (see examples 23 and 24 on page 142). Four different sentence types per filler structure were used: both SVO and OVS sentences with and without plural subject, as well as subject and object relative clauses with and without plural RC subject.

Every sentence in the experiment, including fillers, was 8 words long, and always the second to last word was replaced by an underscore that indicated the gap participants had to fill. Across critical and filler sentences, subjects were required to complete gaps with either an adjective or a full verb in the filler items, or a copula (auxiliary) verb for the critical items. In general, if a subject noticed the number manipulation in the material, what he or she experienced was variation of the number of different NPs across three different structures.⁴

 Procedure

After briefing and screening, subjects were seated in front of a 19-inch LCD screen in a dimly lit sound attenuated booth. They were instructed to attend to a fixation cross on the screen and to a sentence containing a gap, which would be presented in a word-by-word fashion. Subjects were asked to read the sentence, wait for an answer prompt after the complete presentation and then speak aloud the sentence they had read, including their completion of the missing word.

After a 1000 ms fixation cross the sentence appeared in a word-by-word fashion, containing a slide with underscores marking the gap. The end of the sentence was indicated by a period after the last word. Following the final word a visual mask appeared for 500 ms, consisting of a row of hash signs (#). The mask was followed by the production prompt, a single question mark at the centre of the screen. The prompt lasted for 4500 ms, during which subjects had been instructed to respond. After an ITI of 1500 ms the next trial started. The exact timing of an individual trial is shown in figure 7.2 on page 148.

⁴Upon debriefing, all subjects reported that they had not been aware of the experiment’s goal to elicit attraction errors, even if they had noticed the different number markings.
7. Study 3: Production of number agreement

Figure 7.2.: Trial scheme for RSVP stimulus presentation. Vertical arrows indicate the respective timing of noise presentation and answer recording.

In the noise conditions a pre-recorded distractor noise sound was presented. The distractor sound started after the sixth word of a sentence and lasted until the end of the trial, crucially including the time slot allotted to the participants’ response. This way it was made sure that there was little noise interference with reading and memorising the sentence preamble, while still a good portion of the sentence planning and all of the actual articulation of the subjects’ answers took place in noise.

Audiovisual stimulus presentation and sound recording was carried out with E-Prime Pro 2 software (W. Schneider, Eschman & Zuccolotto, 2010). For distractor sound presentation, I used two Genelec 8030A active near-field speakers. The spoken responses of the participants were recorded with a low-latency sound adaptor (ECHO Audio Gina 3G) and an AKG C1000-S microphone, both attached to the presentation PC. The microphone used had a cardioid recording characteristic and was directed at the speaker’s mouth, so I could achieve a good signal-to-noise ratio of the recordings that would allow later scoring of the sentences for all noise conditions. Before the study, the experimental set-up was calibrated with a high-precision sound pressure level meter (Brüel & Kjær Investigator 2260). This way I determined fixed hard- and software settings with which I would reach an average (root mean square, RMS) intensity for the different noise files of 68
7.3. Experiment 3: Agreement error elicitation

dB SPL at the approximate position of a participant’s head.

Apart from silence, which served as baseline condition, three different noise signals were used as distractor stimuli. The signals are part of the International Collegium of Rehabilitative Audiology (ICRA) set of standardised noises for audiological research, simulating spectral and or temporal characteristics of human speech (Dreschler et al., 2001). I chose one unmodulated signal (ICRA 1), with long-term spectrum characteristics of speech (‘pink’ noise), a modulated signal simulating the prosodic contour of one speaker at small distance (ICRA 4), and a signal simulating six talkers speaking at the same time (ICRA 7); see figure 4.1 on page 76 for an illustration of the different kinds of noise. The ICRA 4 signal was used because experiments on language perception have found large negative effects on intelligibility with speech masked by this type of fluctuating sound (Wagener et al., 2006). I included the ICRA 7 ‘babble’ sound as well, because it is conceptually closest to a “cocktail party” setting (Bronkhorst, 2000). For the actual presentation of the different noise sounds I used one 20-second snippets from each of the much longer original sound files. The noise conditions were blocked in order to reduce surprise effects. Block order was counterbalanced across participants and each participant received a different block order on each of the two experimental sessions.

Scoring

Only answers that conformed to the intended structure (see example 21 on page 142) were analysed. The subjects’ answers were scored manually based on the recordings, registering the produced number of both head and local noun as well as the number marked on the verb. An answer was scored as containing an agreement error if the respective number markings on subject head noun and verb did not match. Ungrammatical structures, sentences that did not conform to the intended structure, or unintelligible answers were scored as other. Instances where a subject had produced a different combination of head and local noun number than the one presented were scored as repetition errors, since I could not exclude the possibility that participants had misread or had incorrectly remembered the visually presented sentences. Further analyses of agreement errors were performed based on a data set from which other and repetition errors had been removed; analyses of repetition errors are based on a data set excluding other
errors. Table 7.2 on page 150 provides an overview of the different kinds of errors produced by participants.

Table 7.2.: Error counts by condition, total number of recorded answers $N = 3840$. For each combination of noise condition, head noun number and match condition, 240 answers were recorded. The percentages in figure 7.3 were calculated by dividing the number of agreement errors by the total amount of answers per cell, excluding ‘other’ answers and repetition errors.

<table>
<thead>
<tr>
<th>Noise condition</th>
<th>Error type</th>
<th>Match condition</th>
<th>Mismatch</th>
<th>Match plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Head noun no.</td>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agreement</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Repetition</td>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>silence</td>
<td>Agreement</td>
<td>23</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Repetition</td>
<td>47</td>
<td>53</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>ICRA 1</td>
<td>Agreement</td>
<td>16</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Repetition</td>
<td>37</td>
<td>43</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>10</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>ICRA 4</td>
<td>Agreement</td>
<td>16</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Repetition</td>
<td>36</td>
<td>52</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>14</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>ICRA 7</td>
<td>Agreement</td>
<td>79</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Repetition</td>
<td>163</td>
<td>193</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>45</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Col. sums</td>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Repetition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition to the error counts I also collected data about the time course of the subjects’ articulation. Each response had been recorded separately during the experiment, with the recording onset synchronised to the answer prompt. I gauged the utterance latency by manually measuring from the start of the recording to the onset of a subject’s utterance, using the PRAAT software (Boersma & Weenink, 2011). Cases where a subject had begun speaking before the recording started were excluded. In order to measure speech rate, I also identified the onset of the critical verb in a subject’s answer, and calculated the duration of the initial portion of each ut-
terance. The length of this fragment should reflect the time course of the relevant planning processes up to the point where the articulation of the verb starts. Detailed orthographic transcriptions of the answers were made and served as the data base to count the number of syllables produced in each utterance up to the verb, in order to establish a speech rate measure in syllables per second.

### Design and Analysis

Data from the two sessions per participant were pooled. The data were analysed using mixed effects models (Baayen, 2008; Baayen et al., 2008; Jaeger, 2008). To analyse the different error counts I computed generalised (logistic) mixed effects models with the respective error as outcome variable, and noise, match and head noun phrase number as predictors. Linear mixed effects models were computed for the continuous utterance latency and speech rate data. In both cases, subjects and stimulus sentences were treated as random effects by adding random intercepts to the models. To initially test the results from the baseline silence condition for a mismatch effect, a simple model was specified, based on previous results in the literature, with match condition and head noun number as the only predictors. The same model was then fitted to the entire data set from the experiment, first marginalising over noise conditions. Using the `anova()` function in the R software package (R Development Core Team, 2010; cf. also Baayen et al., 2008), an implementation of a log-likelihood ratio test comparing (Laplace) quasi log-likelihoods, I added noise as a factor and tested for possible interactions with noise by comparing it to the simpler model, as an omnibus test for the factor noise. P-values for linear mixed effects models were calculated with the `pvals.fnc()` function (Baayen et al., 2008).

### 7.3.2. Results

Figure 7.3 on page 152 summarises the number of agreement errors as a function of the match and head NP number conditions in the four different noise conditions. When looking at the data from the silent condition only, it can be seen that the amount of errors is considerably higher in the singular mismatch condition compared to the other conditions. Statistically, the interaction between head noun number and match condition did not reach
7. Study 3: Production of number agreement

Figure 7.3.: Error percentage under different noise conditions.
7.3. Experiment 3: Agreement error elicitation

significance, while the effect of the factor match was significant (see model summary in Table 7.3 on page 153). However, applying the same model to the entire data set, marginalising over noise condition, yields a highly significant interaction effect between the match factor and head noun number (see Table 7.4 on page 153).

Table 7.3: GLMM coefficients for the agreement error data from the silent baseline condition, predictors are treatment-coded (N = 748; log-likelihood = -133.7).

<table>
<thead>
<tr>
<th>Coef</th>
<th>SE(β)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.93</td>
<td>0.51</td>
<td>-7.8</td>
</tr>
<tr>
<td>Head number:PL</td>
<td>-0.52</td>
<td>0.81</td>
<td>-0.6</td>
</tr>
<tr>
<td>Match:MISMATCH</td>
<td>1.71</td>
<td>0.54</td>
<td>3.1</td>
</tr>
<tr>
<td>Hd no.:PL × Match:MISMATCH</td>
<td>-1.33</td>
<td>1.00</td>
<td>-1.3</td>
</tr>
</tbody>
</table>

Table 7.4: GLMM coefficients for the agreement error data for the full data set (predictors are treatment-coded; N = 3026; log-likelihood = -576.2).

<table>
<thead>
<tr>
<th>Coef</th>
<th>SE(β)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.16</td>
<td>0.36</td>
<td>-11.7</td>
</tr>
<tr>
<td>Head number:PL</td>
<td>0.87</td>
<td>0.40</td>
<td>2.2</td>
</tr>
<tr>
<td>Match:MISMATCH</td>
<td>1.62</td>
<td>0.38</td>
<td>4.2</td>
</tr>
<tr>
<td>Hd no.:PL × Match:MISMATCH</td>
<td>-2.36</td>
<td>0.55</td>
<td>-4.3</td>
</tr>
</tbody>
</table>

As described in the previous section, I fitted different models with and without noise condition as a predictor, in order to assess the influence of noise on agreement error incidence. As a result of the model comparison, it turned out that adding noise condition as a predictor significantly improved the model fit (χ²(9) = 18.271, p < .05). I see this as an indication that noise has an effect on the distribution of errors.

Crucially, two interactions of error likelihood with noise reached significance in the final model: When either of the two fluctuating noises was present, the amount of errors increased in sentences with either of the two
7. Study 3: Production of number agreement

nouns marked for plural. Constant noise (ICRA 1) appears to not affect the error distribution. 3-way interactions between noise, match and subject noun number did not improve model fit and were removed from the specification. Table 7.5 on page 154 summarises the final model including noise as a factor.

Table 7.5.: Coefficients for GLM model with noise as factor, predictors are treatment-coded ($N = 3703$; log-likelihood = -761).

<table>
<thead>
<tr>
<th>Coef $\beta$</th>
<th>SE($\beta$)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.28</td>
<td>0.50</td>
<td>-8.5</td>
</tr>
<tr>
<td>Head number:pl</td>
<td>0.03</td>
<td>0.59</td>
<td>-0.1</td>
</tr>
<tr>
<td>Match:mismatch</td>
<td>2.08</td>
<td>0.53</td>
<td>3.9</td>
</tr>
<tr>
<td>Noise:icra1</td>
<td>-0.08</td>
<td>0.56</td>
<td>-0.1</td>
</tr>
<tr>
<td>Noise:icra4</td>
<td>0.25</td>
<td>0.53</td>
<td>0.5</td>
</tr>
<tr>
<td>Noise:icra7</td>
<td>0.22</td>
<td>0.53</td>
<td>0.4</td>
</tr>
<tr>
<td>Hd no:pl x Match:mism.</td>
<td>-2.24</td>
<td>0.57</td>
<td>-3.9</td>
</tr>
<tr>
<td>Hd no:pl x Noise:icra1</td>
<td>0.50</td>
<td>0.60</td>
<td>0.8</td>
</tr>
<tr>
<td>Hd no:pl x Noise:icra4</td>
<td>1.29</td>
<td>0.57</td>
<td>2.3</td>
</tr>
<tr>
<td>Hd no:pl x Noise:icra7</td>
<td>1.21</td>
<td>0.57</td>
<td>2.1</td>
</tr>
<tr>
<td>Match:mism. x Noise:icra1</td>
<td>-0.18</td>
<td>0.59</td>
<td>-0.3</td>
</tr>
<tr>
<td>Match:mism. x Noise:icra4</td>
<td>-0.89</td>
<td>0.56</td>
<td>-1.6</td>
</tr>
<tr>
<td>Match:mism. x Noise:icra7</td>
<td>-0.77</td>
<td>0.56</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

In the analysis of repetition errors, I did not find statistical evidence that noise would significantly improve model fit ($\chi^2(3) = 0.912$, $p = .823$). The baseline model shows a significant effect of head noun number condition, as well as a marginally significant effect of match condition, indicating that conditions which contain a plural marking are more prone to repetition errors than the singular match condition (see table 7.6 on page 155 for a model summary). The distribution of other responses in the data did not appear to be influenced by noise or the different number markings in the sentence preamble.

In contrast, the timing of subjects’ utterances did appear to be influenced by noise. Table 7.7 on page 155 contains mean results of the measurements of utterance latency and speech rate by condition. The model comparison results indicate that the addition of noise as a factor significantly improves model fit for both utterance latency ($\chi^2(3) = 42.69$, $p < .001$) and speech rate ($\chi^2(3) = 29.687$, $p < .001$). Tables 7.8 and 7.9 summarise the final models. No
interaction between the match and head noun number conditions and any of the noise conditions reached significance. Simple effects for each of the three noise conditions indicate that subjects show a lower latency and at the same time a lower speech rate when speaking under noise.

Table 7.7.: Mean utterance latency (in seconds) and mean speech rate (syllables per second) by condition. Total number of analysed recordings $N = 2922$.

<table>
<thead>
<tr>
<th>Noise condition</th>
<th>Measure</th>
<th>Match condition</th>
<th>Mismatch</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Head noun no.</td>
<td></td>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>silence</td>
<td>Latency</td>
<td>.5956</td>
<td>.5958</td>
<td>.5842</td>
</tr>
<tr>
<td></td>
<td>Speech rate</td>
<td>5.898</td>
<td>6.041</td>
<td>5.932</td>
</tr>
<tr>
<td>ICRA 1</td>
<td>Latency</td>
<td>.5511</td>
<td>.5176</td>
<td>.5604</td>
</tr>
<tr>
<td></td>
<td>Speech rate</td>
<td>5.778</td>
<td>5.817</td>
<td>5.876</td>
</tr>
<tr>
<td>ICRA 4</td>
<td>Latency</td>
<td>.5300</td>
<td>.5019</td>
<td>.5301</td>
</tr>
<tr>
<td></td>
<td>Speech rate</td>
<td>5.700</td>
<td>5.716</td>
<td>5.866</td>
</tr>
<tr>
<td>ICRA 7</td>
<td>Latency</td>
<td>.5487</td>
<td>.5475</td>
<td>.5424</td>
</tr>
<tr>
<td></td>
<td>Speech rate</td>
<td>5.840</td>
<td>5.857</td>
<td>5.852</td>
</tr>
<tr>
<td>Col. avgs.</td>
<td>Latency</td>
<td>.5565</td>
<td>.5412</td>
<td>.5548</td>
</tr>
<tr>
<td></td>
<td>Speech rate</td>
<td>5.804</td>
<td>5.859</td>
<td>5.882</td>
</tr>
</tbody>
</table>
7. Study 3: Production of number agreement

Table 7.8.: LME model summary for utterance latency; \( N = 2887 \); log-likelihood = 675.9

<table>
<thead>
<tr>
<th></th>
<th>Coef</th>
<th>SE (β)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.590</td>
<td>0.02</td>
<td>24.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Head number:PL</td>
<td>-0.016</td>
<td>0.01</td>
<td>-1.19</td>
<td>.237</td>
</tr>
<tr>
<td>Match:MISMATCH</td>
<td>0.003</td>
<td>0.01</td>
<td>0.19</td>
<td>.849</td>
</tr>
<tr>
<td>Noise:ICRA1</td>
<td>-0.045</td>
<td>0.01</td>
<td>-4.70</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Noise:ICRA4</td>
<td>-0.061</td>
<td>0.01</td>
<td>-6.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>-0.042</td>
<td>0.01</td>
<td>-4.26</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hd no.:PL x Match:MISM.</td>
<td>-0.003</td>
<td>0.02</td>
<td>-0.13</td>
<td>.886</td>
</tr>
</tbody>
</table>

Table 7.9.: LME model summary for speech rate; \( N = 2920 \); log-likelihood = -3148

<table>
<thead>
<tr>
<th></th>
<th>Coef</th>
<th>SE (β)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.985</td>
<td>0.13</td>
<td>45.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Head number:PL</td>
<td>0.054</td>
<td>0.12</td>
<td>0.45</td>
<td>.605</td>
</tr>
<tr>
<td>Match:MISMATCH</td>
<td>-0.079</td>
<td>0.12</td>
<td>-0.66</td>
<td>.456</td>
</tr>
<tr>
<td>Noise:ICRA1</td>
<td>-0.146</td>
<td>0.04</td>
<td>-4.13</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Noise:ICRA4</td>
<td>-0.183</td>
<td>0.04</td>
<td>-5.16</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Noise:ICRA7</td>
<td>-0.105</td>
<td>0.04</td>
<td>-2.96</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Hd no.:PL x Match:MISM.</td>
<td>0.051</td>
<td>0.17</td>
<td>0.30</td>
<td>.746</td>
</tr>
</tbody>
</table>

7.4. Discussion

7.4.1. Pre-test: Written sentence completion

The results of the pre-test are in line with earlier results obtained by for instance Hemforth and Konieczny (2003) or Fayol et al. (1994) in experiments testing written production. What is striking is the higher amount of errors in both conditions where the subject noun is marked plural. This is in contrast to studies on spoken production, based on which I would not have expected high error rates in conditions with a plural subject and singular local noun (a ‘plural mismatch effect’, cf. Bock & Miller, 1991). Hemforth
and Konieczny (2003) interpret their similar finding as a kind of bias to produce singular verb forms in written production—since in cases where the head noun is plural, such a ‘default’ singular would yield an error (cf. the discussion of morphological complexity in Section 3.4.3). This interpretation would fit in with the analysis of Eberhard (1997), who assumes plural number to be privatively specified (cf. also Wagers et al., 2009 for discussion). However, the singular bias explanation is actually difficult to reconcile with the attraction effect that leads to a higher number of errors in the condition with a plural attractor and singular head noun. If a local noun with plural marking ‘attracts’ plural marking on the verb, I would expect the number of errors in the all plural condition to be markedly lower. Again, this phenomenon might be related to (language-specific) morphological complexity of the inflectional paradigm, see Section 3.4.3. I will return to this question below in the discussion. Finally, since the written production tasks reported by Fayol et al. (1994); Hemforth and Konieczny (2003) and in this article involved reading and did not provide controls for how accurately subjects read the stimulus sentences, I cannot exclude the possibility that the subjects represented number incorrectly on the conceptual level to begin with.

7.4.2. Experiment: Agreement error production in noise

The analysis of agreement attraction effects for the entire data set shows a clear increase in the amount of errors made in the singular mismatch condition (sg-pl), including a significant interaction effect between match condition and head noun number condition. This characteristic interaction, which is expected based on earlier studies, fails to reach significance in the silence baseline condition, however. At present, I have no clear explanation for this, but statistical power issues might play a role. The complete data set shows the characteristic asymmetry of errors occurring more frequently after a plural attractor noun, which replicates the plural markedness effect in attraction (Bock & Miller, 1991; Eberhard, 1997; Hartsuiker et al., 2003).

Crucially, the omnibus test for an effect of noise on the pattern of agreement errors indicates that subjects were influenced by the noise manipulation. This result speaks in favour of the first hypothesis that external noise can influence the sentence production process by creating a secondary task load. The results of the on-line measurements of speaking rate also warrant
this interpretation: under noise, speakers articulate more slowly than in silence. The same effect has been observed before by Postma and Noordanus (1996), and provides strong evidence that speech-free distractor noise can cause interference that results in a slowing down of processing.

The observations about the influence of noise on utterance latency seem a little counter-intuitive, indicating that subjects started their utterances faster with background noise present. A similar observation has been made in study 2 (see Chapter 6). Different causes might play a role for this effect, for instance fluctuations in the arousal level of subjects, or strategies to alleviate the burden on memory imposed by the task. I will return to these possibilities in the general discussion.

An interesting outcome from the analysis of agreement error counts is the fact that only fluctuating noise conditions entered significant interactions with other factors, while I found no indication for constant noise to have a measurable effect on the distribution of agreement errors. This finding might be an indication that the different kinds of noise used in the experiment differ in the severity of disturbance they cause during language production. Such an outcome would actually be expected based on results from the literature on the irrelevant sound effect (e.g. Klatte et al., 1995), where it has been suggested that detrimental effects on performance result from interference by changing-state noise, but not as strongly from interferences by constant noise. As the coefficients from the final model with noise as a factor indicate (see table 7.5 on page 154), the distribution of errors across the Match by Number conditions changes. Error rates increased significantly under fluctuating noise, whenever the head noun was plural.

The plural markedness effect means that under fluctuating noise subjects actually produced more verbs wrongly marked for singular after a plural head. This is surprising especially for the plural match (pl-pl) condition, where the plural markings on both the head noun and the attractor do not provide conflicting cues for assigning the correct number to the verb representation. If a plural local noun is assumed to exert an attraction effect, a spurious number assignment in the plural match condition, based on the local noun’s number marking, should result in (coincidentally) correct number marking.

However, the plural effect I observed might be the result of an increase in ‘default’ singular number markings – an explanation that has been proposed in earlier studies, for instance by Hemforth and Konieczny (2003) or
Franck et al. (2004). This argument gains indirect support by the analysis of the distribution of repetition errors in the data, that is the amount of errors where subjects reproduced the sentence preamble with one or more incorrect number markings on the noun. While I did not find an indication that repetition errors were contingent on the noise condition, there is a statistically significant tendency for this kind of error to appear more often in conditions where more plural markings have to be recognised and/or remembered.

With respect to the agreement error data, how could (fluctuating) noise lead to a stronger reliance of subjects on the default or unmarked form? A first step towards an explanation might be provided by looking at models for the irrelevant sound effect (ISE). In particular, the changing state hypothesis by Jones and Macken (1993) and Macken et al. (1999) emphasises the role of the structured nature of both speech as well as fluctuating, non-speech distractor sounds (see Chapter 4). The authors suggest that any structured sound signal is automatically subjected to auditory scene analysis, which tries to extract an ordered sequence of auditory objects from the signal. The hypothesis predicts that processing for the primary task of maintaining an ordered list of words in working memory will compete for processing resources on a seriation process, that would be engaged concurrently by automatic auditory scene analysis.

As noted in Chapter 4, the task of producing sentences with correct agreement cannot be straightforwardly compared to the serial verbal recall tasks usually employed to investigate irrelevant sound effects. However, maintenance of the order of elements might be a basic function that serves on-

---

5 In the context of our present study, a superficial check of lexical frequencies of the verb forms of the German verb ‘sein’ (to be) with the Leipziger Wortschatz database reveals that the 3rd person singular form ‘ist’ is at least one order of magnitude more frequent than the respective plural form ‘sind’. This is also confirmed by frequency counts from the CELEX database (Baayen, Piepenbrock & Gulikers, 1995; Max Planck Institute for Psycholinguistics, 2001). Based on frequency data from the CELEX corpus, I used the formula suggested by Kostić (1991) to calculate a rough entropy estimate for the different word forms of the auxiliary verb ‘sein’ (to be), which was used by almost all of the experimental subjects for the completion of test sentences. My rough estimate for information content provides further evidence for a ‘default’ status of the 3rd person singular form of the verb ‘sein’ over the respective plural form. Hence, the higher number of singular forms produced under noise can be related to the default status of singular, be it in terms of the word form frequency or in terms of informational content or complexity of the singular form.
7. Study 3: Production of number agreement

line sentence generation and the computation of agreement at some point during processing. Competition for processing time on a seriation mechanism should lead to a slowing down of concurrent language processing. If under (fluctuating) noise processing is slowed down, retrieval of the agreement controller noun phrase can in some cases take too long, and an empty or ‘default’ number specification is used as a cue for the retrieval of the verb form. Alternatively, the plural cue might be set too late during verb form retrieval, so that the default verb form will win by means of a head start in activation because of its higher frequency. It should be possible to integrate both accounts with different existing models for verb agreement, for instance the Marking and Morphing model by Bock, Eberhard and Cutting (2004), or the working memory retrieval model outlined in Badecker and Kuminiak (2007; also cf. Lewis & Vasishth, 2005 and Wagers et al., 2009 and see Section 3.4.3 in Chapter 3). Whatever the precise mechanism, the effect of fluctuating noise I observed is not an effect of attraction as such, but rather follows from the defaulting to singular verb forms, which are inappropriate after a plural head noun. To put it differently, I might say that under the difficult conditions no agreement relation is established at all between a verb and its controller.

⁶This might also serve as a cornerstone for an explanation of the unexpected results from the pl-pl condition in the written sentence completion task: The activation of the plural feature spawned by the plural distractor might lead to higher uncertainty about the correct number of the actual agreement controller. A possible way to cope with this uncertainty is to resort to a ‘default’ strategy.
Part III.

Discussion
8. General discussion

As stated in the introduction to this thesis, the goal of the research project reported here was to charter the ground for the question how acoustically difficult circumstances influence the generation of complex sentence structure. With the theoretical Chapters 2 through 4 I provided the foundation for a systematic investigation of different linguistic structures and identified some necessary cornerstones for modelling the process of producing language under noise in greater detail. Chapters 5 through 7 reported the data collection with which I intended to ‘zoom in’ on the phenomenon.

After addressing a couple of points on the methods used for the empirical studies, the following discussion will relate the study results reported in part two to the theoretical issues raised in part one. Because of the exploratory nature of this work I will have to exert some caution with respect to the generalisations drawn from my data. For the same reason a comprehensive model for the production process under noise is beyond the scope of this thesis. I will however point to a number of interesting boundary conditions for designing a language production model that I see to be falling from my results.

8.1. Methodological challenges

8.1.1. Ecological validity and statistical power

In comparison to experimental studies on language comprehension, experiments on language production are quite rare. Part of the reason might be that production studies face a number of inherent methodological challenges, for instance being to a large extent dependent on the outcome of the language production process, rather than being able to observe processing on-line.

Corpus analyses like the one conducted for study 1 have the advantage of tapping the most ‘natural’ production processes, because the assembly of
8. General discussion

A corpus can take place under comparatively non-artificial circumstances. The disadvantage is the large amount of data necessary to gain statistically valid insights. For study 1, Michael Vitevich (pers. comm.) commented on the small number of participants, and pointed to possible power issues with respect to the statistical methods used for study 1. In sum, the setting chosen was the result of a general problem or trade-off between statistical power and a feasible experimental design, given that the scoring method required considerable amounts of manual transcription and analysis.

The more controlled experiments fare better in this respect because they included a larger number of subjects from which sentences were sampled. However, the greater control comes at the price of creating an artificial situation, which necessarily limits the possibility for easily generalising results to natural communication situations. For example, it might be argued that the controlled experimental conditions in studies 2 and 3 do not pose any restrictions on the information structure of a sentence that is about to be produced. With respect to the observations made in study 2, certain ‘natural’ contexts might have made object relative clauses the preferred structure over subject relative clauses or relative clauses with passive, see for instance Mak et al. (2008) or Weskott (2002) for evidence from comprehension studies. Weskott argues that information structure conveys a processing order for search of the discourse model in comprehension. During production the object relative clause construction serves an information-structural function, allowing to convey a particular state of affairs represented in a discourse model, which might not be warranted by experimental conditions. Therefore I have to refrain from generalisations based on linguistic context and focus on the perceptual or environmental context, which was actually manipulated in the studies.

8.1.2. Apparatus

All studies have in common the fact that noise was presented through loudspeakers in a well-controlled free-field. Also, the order of magnitude of noise intensity was similar for all studies, between 65 and 68 dB SPL.

Martin Cooke (pers. comm.) voiced concerns that the results of study 1 might not actually show a Lombard effect, due to comparatively low intensity of noise. Other studies researching the Lombard effect and other effects of noise on speech have used noise intensities of up to 100 dB SPL.
8.1. Methodological challenges

(see e. g. Garnier et al., 2006; Junqua et al., 1999; Postma & Noordanus, 1996; Zhao & Jurafsky, 2009). Cooke also demurred on the experimental setting with presentation of noise through loudspeakers. Under these circumstances certain measurements on the recorded speech signal are more difficult or even impossible, because the recording necessarily contains portions of the noise as well. My response to this concern is that the choice of apparatus also resulted from a trade-off between measurement accuracy and higher ecological validity of the setting. In anticipation of criticism like Cooke’s, great care was exercised with the selection of appropriate microphone hardware, and an appropriate formula was used for estimating the signal-to-noise difference for voice in noise.

Nevertheless, while the Lombard effect does not form the focus of this thesis, both concerns have to be taken seriously, and I cannot exclude the possibility that the increase in voice intensity reported in study 1 is the result of a measurement artefact. For the purpose of an exploratory study, however, the setting offered good enough results in that it yielded at least some indication that noise has an effect on the speakers.

8.1.3. Age effects

The experiments carried out for study 2 compared the effects of noise on language production in subjects both with and without hearing impairment, and across a wide age range. The results indicate that among subjects with hearing impairment, noise and perceptual difficulty might result in cognitive costs which impair the smooth processing of complex sentences.

However, the age effects observable in the data give a somewhat mixed picture and are not fully consistent with other studies on cognitive ageing, for instance by Susan Kemper and colleagues. One reason might be that the design of the study reported here differed in some crucial methodological points from others: First of all, I chose to test a mid-aged group of subjects, between 29 and 59 years old, and participants in this group might not yet show typical age-related effects of cognitive decline. What is more, the actual composition of the subject groups formed to compare speakers with hearing impairment to speakers with normal hearing was such that age acts as a confounding variable. The higher-order interactions between hearing status (a dichotomous variable) and age (a continuous variable) render the results difficult to interpret, and any causal interpretation of the correla-
8. General discussion

...tions between hearing status and age have to be taken with a grain of salt as statistical measures do not allow for a post-hoc control of group differences in age (see Miller & Chapman, 2001).

8.2. Summary of the empirical results

8.2.1. Effects on structural complexity

The first study (Chapter 5) analysed a corpus of near-spontaneous speech. The experimental study resulted in two sub-corpora, which contrasted recordings of speech produced in noise and speech produced in silence. The search for tokens of structures deemed ‘complex’ or difficult yielded no significant differences between silence and noise, hence I cannot infer an effect of noise on structural complexity in the utterances produced by the subjects.

With study 2 (Chapter 6) I created an experimental situation in which speakers were asked to produce particular structures that have been considered difficult on theoretical grounds. The study thus tried to (artificially) increase the chance of subjects producing ‘difficult’ structure and to increase the likelihood of an interference between the processing of noise and the sentence generation processes. Experiment 2.1 (Section 6.2) generally yielded evidence for a preference for active over passive sentence structure. This preference appears to be modulated by noise only in subjects with hearing impairment: a higher variance between active and passive sentence structure in silence is reduced to a level more similar to that produced by subjects with normal hearing. Put differently, there might be a small but significant tendency for subjects with hearing impairment to rely on the ‘simpler’ active sentence structure under noise in a priming experiment. With experiment 2.2 (Section 6.3) I presented evidence for a preference for subject relative clauses with passive over object relative clauses. This is in line with earlier observations from the literature (e. g. Belletti, 2009; Contemori & Belletti, in press). The addition of noise appears to have the most pronounced effect on subjects without hearing impairment. Unfortunately, for both experiments the factor hearing status is confounded with age, which makes the intended contrast between subjects with and without hearing impairment difficult to interpret.
8.2. Summary of the empirical results

8.2.2. Effects on the amount of errors

A post-hoc analysis of error tokens from study 1 indicates that grammatical errors might be more prevalent under noise. In particular, agreement errors appear to be more likely, although their numbers are quite small (the number of errors made in the experiments 2.1 and 2.2 was too small to analyse statistically). In order to follow up on this assumption, I designed the error elicitation experiment reported in study 3 (Chapter 7), which investigated the effect of noise on grammatical errors more systematically. The results of experiment 3 show that agreement errors are more likely under fluctuating noise. The pattern of errors made by speakers indicates that subjects relied more on a ‘default’ singular number marking. Crucially, the disturbance by noise did not only lead to an increase in agreement errors, but the choice of a default form seems to indicate that correct agreement is not established at all in some cases.

8.2.3. Effects on utterance latency and speech rate

In order to gauge effects of noise on the production process as it unfolds, I recorded two on-line measures for the speed or time-course of sentence production, utterance latency and speech rate. As the results of experiments 2.1 and 2.2 show, the time needed to initiate an utterance appears to be influenced by the structure a speaker is about to use: in experiment 2.1 the latency for active sentences is smaller than for passive sentences. Crucially, in the same experiment the utterance latency generally decreased under noise. The latency data from experiment 2.2 on the production of relative clause structure show a more complicated pattern, because of a higher-order correlation with age and hearing status. The results indicate that for subjects with hearing impairment, the difference between the noise and silence condition is reduced compared to subjects with normal hearing. On the other hand, the results indicate that at least for normal hearing speakers, similar to what was observed in experiment 2.1, utterance latency decreases under noise. This trend is corroborated by the results of experiment 3, in which I also found faster utterance initiation times under noise.

The effects of noise on speech rate are less homogeneous. In study 1, the presence of noise had no effect on speech rate. This appears to be the case for the experiments of study 2 as well, although the results of the prim-
8. General discussion

ing experiment (2.1) indicate that passive sentences are produced with a higher speech rate than active sentences. In addition, subjects with normal hearing tend to speak faster than subjects with hearing impairment. This appears to be in line with the observation from experiment 2.2 that subjects with hearing impairment speak slower when producing object relative clauses. However, higher-order correlations with age make this result difficult to interpret. Finally, experiment 3 provides some indication that a noise setting results in a slower speech rate compared to silence.

8.3. Interpretation

The results presented in the empirical part of this thesis yield only little obvious effects on the complexity measures for the utterances produced by subjects. Contrary to what would have been expected based on the study by Kemper et al. (2003) the results do not indicate an obvious strategy for subjects to resort to producing ‘simpler’ sentence structure under adverse conditions. Some effects on on-line measures of production time course indicate that sentence formulation does not remain unaffected by noise, however. In my opinion the most convincing evidence is that found in experiment 3, which showed that on a structural level production is more error-prone under noise than in silence.

8.3.1. Strategy

One way to interpret the results on latency and speech rate might be to postulate strategic differences, i.e. changes in processing strategy to overcome any perceptual or environmental difficulty caused by hearing impairment and/or noise. This could be seen as a way for speakers to ‘cope’ with the perceptually adverse situation. Kemper et al. (2003) further argue that older adults “develop a restricted speech register that is grammatically less complex and propositionally less dense than that normally used by young adults” (p. 189). As far as acoustic measures of speech in noise are concerned, like the Lombard effect, (Cooke et al., 2007) also discuss the possibility of strategic adaptions by speakers, but remain uncertain whether those could happen sufficiently fast. The on-line processing measures in study 2 yield some indication that more complex structure prompts small changes in the behaviour of subjects with hearing impairment, who tended
to increase their speaking rate on the matrix clause before object relative clauses and subject relative clauses with passive. Thus a potential generalisation would be to assume that difficulty during speaking in noise can be overcome by changing the type of structure used (which I cannot confirm to be the case based on my data), as well as changing the rate at which we speak.

However, since subjects participating in the studies reported here had not been trained or instructed in any particular direction and only in study 1 were placed in a ‘conversational’ setting, the observed behaviour might also reflect an automatic adaptation to the noisy setting. In addition, it appears unlikely that the increase in error rate observed in studies 1 and 3 is the result of speakers strategically deciding to make more errors—rather, the increase in error rate is more likely the effect of an underlying change in automatic/unconscious processing. The postulation of a strategic difference therefore does not actually explain the changes in speech rate or utterance latency, unless we assume these changes to be under intentional control. Especially with respect to processing ‘preferences’ for certain types of structure (see the discussion of experiment 2.2), there is to my knowledge so far no convincing approach for a mechanistic model of ‘strategic’ choices for one type of structure over another.

### 8.3.2. Working memory mechanisms

Instead of an explanation in terms of strategy, parts of the research carried out for this thesis were intended to aim for a more detailed, mechanistic explanation. Following previous work on the irrelevant sound effect a likely component of such an explanation is working memory (WM). Properties of working memory have been independently established with other experimental paradigms, and include for instance capacity limitation or the susceptibility to interference (also see Chapter 4).

**Capacity limitations**

A recurring line of argumentation is that processing a more difficult type of sentence structure while having to restaurate a distorted speech signal during comprehension will cost processing ‘capacity’ or ‘resources’, and the ensuing competition for resources will negatively affect sentence pro-
cessing. In a similar vein, effects of hearing impairment on language comprehension have been explained using the notion of reduced capacity due to higher “filtering” demands (McCoy et al., 2005; Wingfield et al., 2003). A similar explanation is conceivable for language production, where capacity limitations might disrupt the formulation process.

However, the data from studies 1 and 2 do not show a direct influence of the presence of background noise for normal hearing subjects with off-line measures on sentence complexity. Even with the results on speech rate and utterance latency, I did not find comparable higher-order interactions between perceptual difficulty, stimulus degradation and syntactic complexity like those shown for instance by Wingfield et al. (2006) for comprehension tasks. It is possible that for unimpaired adult speakers the acoustic effects of concurrent background noise are simply not detrimental enough to yield an effect on the choice of sentence structure. Effects might become apparent only in combination with other factors that are taking their toll on processing capacity, for instance hearing impairment or cognitive ageing.

Interesting in this respect is the finding from study 2 that subjects with hearing impairment did show a tendency to resort to simpler structure, in that variation between active and passive is reduced under noise in one condition of the priming experiment. The stronger preference of the subjects with hearing impairment to produce passive subject relative clause structure instead of object relative clauses points in the same direction. This could be seen as an indication that the perceptual difficulty experienced by speakers with hearing impairment creates processing load, which might be alleviated by resorting to less ‘complex’ sentence structure.

There are two issues with this interpretation, however. While for comprehension studies the higher cognitive effort to repair a mangled speech signal is a somewhat plausible source of ‘cognitive load’ or ‘effort’ (Rabbitt, 1968; Wingfield et al., 2006), it remains unclear how such an additional effort should arise during language production. What also seems at odds with this interpretation is that I did not see a similar effect of noise for the passive priming condition. It should be borne in mind, however, that in this condition the passive sentence structure was of higher availability to the subjects. Increased availability of a particular structural type because of priming has been shown earlier for agrammatic subjects which had severe difficulty to produce passive sentences without priming (Hartsuiker & Kolk, 1998).
As I have already argued in Chapter 4, a simple explanation of noise-related sentence production phenomena in terms of unspecific ‘capacity’, similar to some explanations for comprehension phenomena seems to be of limited use, and the notion of ‘capacity’ needs to be more clearly substantiated. Following for instance Rummer et al. (1998), the view taken here is that such a narrower definition should work best in terms of a mechanism and its limitations in terms of processing ‘time’ and/or ‘space’ (Hartsuiker & Kolk, 1998; Salthouse, 1988).

An explanation for inter-individual differences between groups of subjects along the lines of the ‘time’ metaphor has been proposed by Hartsuiker and Kolk (1998) for the language of patients with Broca-type aphasia. The authors have proposed that patients suffer from a timing or synchronization problem within a slowed-down language processor. Taking up this analogy I might argue that for unimpaired, middle-aged adult speakers the temporal fine-tuning is working optimally, hence sufficient processing time resources are available. As for processing ‘space’ restrictions, the passive structures elicited in study 2 incurred a longer utterance latency across participants and conditions. This is in line with previous findings (F. Ferreira, 1991) and provides some indication that the amount of structure prepared during on-line processing plays a role.¹

In an earlier experiment using the agreement error elicitation paradigm that was replicated in study 3, Hartsuiker and Barkhuysen (2006) found evidence for inter-individual differences in the likelihood to make agreement errors that correlate with a sentence-span working memory measure. Based on these observations, the authors claim that it takes WM ‘resources’ to reliably (re-)establish the correct number marking of the agreement controller. However, their study does not distinguish between the two capacity metaphors time and space (although the use of a reading span as the working memory measure correlated with the results of the experimental task at a first glance suggests a space metaphor). The results of study 3 on agreement error production show an effect of noise on the amount of errors, and indicate changes to the time-course of processing under detrimental conditions. Given the general view on working memory and processing adopted

¹As mentioned previously in Chapter 6, the faster speaking rate for passive observed in both experiments is probably linked to the presence of an additional function word in passives as compared to active sentences.
in the theoretical part of this thesis, I take it that the assumption of a capacity limitation is a hypothesis about a processing mechanism. Hence as stated earlier a clear desideratum is to attempt a more explicit modelling of how linguistic processes interact or interfere with properties of our (verbal) working memory, cf. for instance Lewis and Vasishth (2005).

Interference

Interference within memory forms a crucial part of different explanations for agreement attraction effects. The effects observed in study 3, in particular the written sentence completion pre-test replicate results from earlier studies and are compatible with different models proposed in the literature. What is more, as has been argued from a theoretical linguistic perspective, interference effects might also play a role in the apparent preference for a less ‘difficult’ structure like SRC or SRC with passive over ORC as seen in study 2 (Belletti & Chesi, 2011; Contemori & Belletti, in press; Friedmann et al., 2009, also see Mak et al., 2008). Crucially, the phenomena are linked by the fact that the interference is due to similarity between elements of the sentence, i.e. morphemes or words. In more general terms, this kind interference effect might be due to “similarity of content,” as Macken et al. (1999) have called it.

The story for the spoken agreement error elicitation experiment is not that simple however. This is because (a) the subjects showed a ‘default’ strategy rather than selecting the wrong, interfering agreement controller, and because (b) the influence of speech-free noise is difficult to conceptualise as a case of content similariy—the distracter noise used in studies 1 through 3 was specifically chosen to contain no linguistic information.

Dual task-interference and the ISE

An alternative explanation thus might be to assume interference by “similarity of process” (Macken et al., 1999). As described in Chapter 4, this hypothesis postulates that there is competition for shared functions used in both processing of noise as well as processing-for-production. In turn, this view implies that in cases where a secondary task load has a detrimental effect on language processing, the secondary task makes use of resources,
8.3. Interpretation

that is it employs basic procedures, which are also used for the language-related task, and competition ensues (cf. e. g. Berti, 2010; Caplan & Waters, 1999)—also see the discussion in Chapter 4.

More specifically, the changing state hypothesis by Macken and colleagues helps to explain the finding from study 3 that ‘babble’ noise has the most profound effect on the generation of agreement: The automatic processing of background noise (Jones & Macken, 1993; Jones et al., 1992) requires ‘parsing’ functions (see Chapter 4). The changing state information in the babble noise might be seen to form a kind of artificial prosody, which could prompt the language comprehension system to search for prosodic boundaries, in order to form chunks of information from the sound stream. Keeping track of the temporal order of chunks in turn is a function that could be very likely required during the language production process.

Decay

Another crucial characteristic of the memory system as specified by Lewis and Vasishth (2005), which in turn is based on the ACT-R framework (Anderson, 2007), is the decay of an item or chunk’s activation as a function of time. As in many other models using the metaphor of a network with spreading activation, decay is a fundamental property which helps to capture phenomena like the limited focus of attention and the subsequent necessity to reactivate chunks from memory. The assumption of decay of (parts of) data chunks in memory forms an important step in the explanation of the agreement error data observed in experiment 3. As indicated earlier in the discussion of study 3 and in this general discussion, interference from an irrelevant noise effect might slow down processing, because the automatic parsing of noise competes for a serial ordering function. Lewis and Vasishth (2005) and Badecker and Kuminiak (2007) envision language processing for both comprehension and production to require constant retrieval of information from immediate memory. This retrieval is based on formal cues, which are also subject to decay over time. As a result of the automatic activation decay of the cues necessary for retrieving earlier elements in a sentence structure representation from memory, speakers practically ‘forget’ what the agreement controller’s number was. Because number information is not available any more at the verb site, the ‘default’ number is chosen for the verb form retrieval/assembly. Which of
8. General discussion

the inflected verb forms in a paradigm serves as ‘default’ in a particular language falls from distributional properties of morphological paradigms and can be captured by general mathematical concepts like entropy (see Chapter 3 and the discussion of study 3 in Chapter 7).

Finally, further indication for a timing problem comes from the observation that utterance latency is reduced while speaking in noise. The latency reduction might reflect a kind of ‘strategy’ to alleviate the problem of decaying memory representations. As of now, it remains an open question whether this kind of strategy is pursued intentionally or is the symptom of an automatic regulation process (see discussion in Section 8.3.1).

8.3.3. The role of the monitor

After decomposing detrimental effects of noise on language production into individual working memory functions I would like to take a step back again and connect these mechanistic assumptions with more general concepts about language production, in particular with the monitor (cf. Chapter 3). The monitoring system as envisioned by Levelt (1983, 1989) has faced some criticism in recent years. Postma (2000), Hartsuiker and Kolk (2001), and Vigliocco and Hartsuiker (2002) question the plausibility of an internal loop that feeds a pre-articulatory phonological plan into the monitor where it is available to conscious inspection, as some kind of ‘internal voice’. Vigliocco and Hartsuiker (2002) rightly point out that this might lead to synchronization problems between the self-perception of external, overt voice and the internal voice, which should lead to the constant perception of an echo. Instead, they and Postma (2000) argue for a different conception of internal monitoring in terms of production-internal feedback, which is monitored by distributed subsystems not necessarily available to conscious inspection (also see MacKay, 1992 and Levelt, 1992 for discussion).

Far less controversial is the assumption of an external loop, which processes speech through the auditory perceptual system. The external loop seems to form a plausible candidate for the necessary interface between perception and production at which external noise might exert an influence. Postma and Noordanus (1996) support this assumption with experimental data on the negative influence of noise on the external loop. According to their results, impairment of the monitor’s proper function via the perceptual loop leads to higher error rates on the phonological and lexical level.
8.3. Interpretation

Possible effects of noise on formulation might well stem from two different sources, (i) the degradation of the monitored own speech, and (ii) an irrelevant sound effect. Similar to what I have proposed as an explanation for the influence of noise on agreement based on Macken et al. (1999), changing-state noise might generate competition for processing time/space on resources shared between the automatic parsing of the distractor noise and monitoring.³ What is more, by masking or degrading his or her own speech, the phonetic or phonological memory representation of what a speaker has said will be impoverished and cannot serve as an additional cue for retrieving the correct verb form during the generation of agreement.

The ‘self-perception’ during speaking through the external and internal loop might thus actually serve an even more general function during speaking than monitoring for errors, it might also serve to reinforce the memory representation of parts of speech that have been produced earlier. Self-perception could help to strengthen search cues to re-access earlier parts of the utterance from immediate memory, at positions when this is necessary to process the current part of the utterance (cf. Badecker & Kuminiak, 2007; Lewis & Vasishth, 2005). If the self-perception through different monitoring loops is impaired in noise, the fact that we observe an increase in errors like wrong agreement marking or interleaved sentences (cf. Chapter 2) could actually be seen as the effect of a less efficient access to parts of the conceptual, grammatical or phonological structure of what one was about to say. What is more, my results about a reduced speech rate in noise, as well as earlier finding to this end (cf. Section 4.3.2 in Chapter 4.4.1) might be explained by recourse to a slowing down of processing in the monitor. The question remains how impaired monitoring through the external loop, which always happens after articulation has already taken place, could lead to a higher number of grammaticality errors (see Chapter 5, in particular Section 5.3.5). The findings of the detailed error analysis from study 1 do not preclude an explanation based on external monitoring, as almost all error types I found could result from a failure to keep track of parts of the sentence produced moments ago. Since the detailed error analysis in study 1 was carried out post hoc, this question warrants further research under

³It is possible, of course, that a competition for processing resources instigated by irrelevant sound also affects the effectiveness of an internal monitoring loop, in addition to the degradation of the perceived signal.
more controlled conditions.

Finally, in their explanation of the Lombard effect or reflex, Lu and Cooke (2008) assume an automatic regulation mechanism underlying the changes in speakers voice under noisy conditions. This mechanism should in principle be compatible with control mechanisms postulated for speech monitoring, or might even form part of the monitor. The Lombard reflex might be different though in terms of the regulatory mechanism involved, which influences muscular control necessary for phonation and articulation. In effect the reflex might nevertheless be in a speaker’s own interest: to reinforce the self-perception during speaking, in order to reduce the likelihood of errors or to even more fundamentally support memory retrieval processes necessary for the construction of sentence structure on-the-fly.
9. Conclusion and outlook

Taken together, the studies reported here speak for an important role of *acoustic proprioception* for language production, through the external or overt feedback loop of the monitoring system proposed by Levelt and others. The representation of what we have said earlier for can yield additional cues for memory retrieval of correctly specified word forms. These findings support the assumption that the monitoring mechanism is an integral part of the host of processes that establish syntactic structure (cf. Franck et al., 2010, whose analysis of agreement errors from a more grammar-theoretical perspective also invokes the notion of a monitor).

Language comprehension has been shown in earlier work to be not only bottom-up or data driven, but also top-down, driven by expectations and context information to some extent. Similarly, the results reported here lead me to believe that language production is not only top-down, but bottom-up, or partially perception-driven. However, the production process seems to be not as susceptible to external influence as one might intuitively think. Core structure assembly processes which allow us to construct ‘difficult’ sentence structure appear to be more robust, or in other words more automatic/informationally encapsulated/domain-specific or functionally specialised than others. It turns out that listening to noise as an unrelated secondary task does not overlap far enough with highly specialised routines for sentence structure assembly to disturb them in an experiment.

Based on the results of studies 1 and 3 it is argued here that the monitor is susceptible to interference. The agreement attraction data does not allow for a straightforward decision between different processing models for agreement production, but the effect I observed might add to the empirical base on which different models are being constructed. Importantly, the finding that something as common or unspecific as background noise can influence grammatical encoding speaks against the assumption that the formulation stage operates in a highly automatic fashion with dedicated, or ‘exclusive’ processing resources at its disposal (V. S. Ferreira & Pashler,
9. Conclusion and outlook

2002; Garrod & Pickering, 2007; Hartsuiker & Barkhuysen, 2006). Rather, as other authors have argued before, automaticity appears to be a matter of degree.

The exploratory work pursued for this thesis forms an important step to further investigate the effects of noise on language production. As the subject matter has proven to be complex and difficult to investigate empirically, more strictly controlled experimental tasks should follow, which might increase the burden on the speakers’ processing capacity. In addition, based on the observations made in experiment 3, a closer examination of the effects of different kinds of noise at different intensity levels should be useful for further detailing the properties of processes shared between language production and other cognitive tasks. This way it should for instance be possible to further test the hypothesis that language production processes compete for processing time on a domain-general ‘seriation mechanism’.

More generally, I have argued for the view that processing under dual-task conditions can lead to interference between basic processing steps or ‘functions’ that are shared between cognitive tasks. This view presents exciting opportunities to further investigate the complex processing mechanisms involved in production (and comprehension) of language, in order to dissociate domain-general and domain-specific processing functions (cf. for instance Barrett & Kurzban, 2006). Using different measures for cognitive ‘capacity’, i.e. employing different secondary tasks might provides us with a valuable tool to more closely characterise the mechanics of language processing. Going from here, more candidates for perceptual or general cognitive processes that serve language processing should be identified and tested.

Following the work by researchers like Marr or Jackendoff, I have argued that a psychologically plausible model for linguistic cognition might be attainable (a) by aiming for a clear distinction between levels of explanation in Marr’s sense and (b) by taking every effort to ‘de-construct’ domain-specific principles and to reduce them to principles of cognition commensurate with theories from within and without one’s own discipline. As I have argued in my discussion, independently established, domain-general properties of working memory like decay and similarity-based interference of content and processes can provide us with the core elements for an explanation of domain-specific phenomena.
A. Stimulus material study 1

A.1. Interview questions: Holidays

In this part of the interview I would like to talk with you about the ‘major’ holidays, like Easter, Pentecost, or Christmas.¹

1. Can you remember the last big holiday occasion(s)?
2. Do you and your family have any particular rituals for celebrating larger holidays?
3. Are there special meals you would eat on one of those occasions?
4. Do you experience a lot of hustle and bustle over the holidays?
5. How would you cope with holiday stress?
6. What was the last present you got for a holiday occasion?
7. If tomorrow were Christmas again and you could make a wish for something, what would that be?

A.2. Interview questions: Travel

In this part of the interview, I would like to talk to you about holiday travel.

1. The last time you went on holiday, where did you travel?
2. How did you travel there?
3. Why did you travel by [helicopter]? ¹

¹I featured the three examples as ‘major’ holidays since they involve official bank holidays all over Germany, and are usually accompanied by at least one week of school vacation in most parts of the country.
A. *Stimulus material study 1*

4. What is your favourite means of transport for going on holiday?

5. What do you like about travelling by *helicopter*?

6. How do you usually prepare for a trip?

7. If you could travel anywhere, where would you like to go?

8. Why would you like to go to *Mars*?
A.3. Short picture story: Well intended

Figure A.1: *Gut gemeint* ('Well intended') by e. o. plauen, from "Vater und Sohn" in *Gesamtausgabe Erich Ohser* (Ohser, 2000), © Südverlag GmbH, Konstanz, 2000.
A. Stimulus material study 1

A.4. Short picture story: Peacemaker

Figure A.2.: Friedensstifter (‘Peacemaker’) by e. o. plauen, from “Vater und Sohn” in Gesamtausgabe Erich Ohser (Ohser, 2000), © Südverlag GmbH, Konstanz, 2000.
B. Stimulus material study 2

B.1. Exp. 2.1: Active/passive sentence priming

Stimulus sentences

Each proposition depicted in the stimulus pictures was realised as an active (A) and a passive (P) sentence.

   “Here the pirate is shooting the soldier.”  
   P: Hier wird der Soldat vom Piraten erschossen.  
   “Here the soldier is being shot by the pirate.”

   “Here the cowboy is catching the soldier.”  
   P: Hier wird der Soldat vom Cowboy gefangen.  
   “Here the soldier is being caught by the cowboy.”

   “Here the waiter is interviewing the hairdresser.”  
   P: Hier wird der Friseur vom Kellner interviewt.  
   “Here the hairdresser is being interviewed by the waiter.”

   “Here the tourist is interviewing the knight.”  
   P: Hier wird der Ritter vom Touristen interviewt.  
   “Here the knight is being interviewed by the tourist.”

   “Here the robber is chasing the magician.”  
   P: Hier wird der Zauberer vom Räuber verfolgt.  
   “Here the magician is being chased by the robber.”

   “Here the tourist is chasing the painter.”  
   P: Hier wird der Maler vom Touristen verfolgt.  
   “Here the painter is being chased by the tourist.”
7. A: Hier erschießt der Polizist die Köchin.
   “Here the policeman is shooting the female chef.”
   P: Hier wird die Köchin vom Polizisten erschossen.
   “Here the female chef is being shot by the policeman.”

   “Here the ghost is filming the magician.”
   P: Hier wird der Zauberer vom Gespenst gefilmt.
   “Here the magician is being filmed by the ghost.”

   “Here the tourist is interviewing the girl.”
   P: Hier wird das Mädchen vom Touristen interviewt.
   “Here the girl is being interviewed by the tourist.”

10. A: Hier verfolgt die Nonne den Räuber.
    “Here the nun is chasing the robber.”
    P: Hier wird der Räuber von der Nonne verfolgt.
    “Here the robber is being chased by the nun.”

    “Here the pope is touching the detectives.”
    P: Hier werden die Detektive vom Papst berührt.
    “Here the detectives are being touched by the pope.”

    “Here the vampires are observing the painter.”
    P: Hier wird der Maler von den Vampiren beobachtet.
    “Here the painter is being observed by the vampires.”

    “Here the chef is shooting the female tourists.”
    P: Hier werden die Touristinnen vom Koch erschossen.
    “Here the female tourists are being shot by the chef.”

    “Here the baker is interviewing the robbers.”
    P: Hier wird der Räuber von den Bäckern interviewt.
    “Here the robbers are being interviewed by the baker.”

15. A: Hier filmt das Monster die Frisörinnen.
    “Here the monster is filming the female hairdressers.”
    P: Hier werden die Friseurinnen von dem Monster gefilmt.
    “Here the female hairdressers are being filmed by the monster.”
B.1. Exp. 2.1: Active/passive sentence priming

16. A: Hier filmen die Omas die Königin.
   “Here the grandmother is filming the queen.”
   P: Hier wird die Königin von den Omas gefilmt.
   “Here the queen is being filmed by the grandmother.”

17. A: Hier grüßt die Ärztin die Zimmermänner.
   “Here the female doctor is greeting the carpenters.”
   P: Hier werden die Zimmermänner von der Ärztin gegrüßt.
   “Here the carpenters are being greeted by the female doctor.”

18. A: Hier grüssen die Bäuerinnen die Ärztin.
   “Here the female farmers are greeting the female doctor.”
   P: Hier wird die Ärztin von den Bäuerinnen gegrüßt.
   “Here the female doctor is being greeted by the female farmers.”

19. A: Hier tadeln die Soldaten die Touristin.
   “Here the soldiers are reprimanding the female tourist.”
   P: Hier wird die Touristin von den Soldaten getadelt.
   “Here the female tourist is being reprimanded by the soldiers.”

20. A: Hier verscheucht die Kellnerin die Feen.
   “Here the waitress is shooing away the fairies.”
   P: Hier werden die Feen von der Kellnerin verscheucht.
   “Here the fairies are being shooed away by the waitress.”
Stimulus pictures¹

1. a) Pirate shooting soldier   b) Soldier shooting pirate

2. a) Cowboy catching soldier   b) Soldier catching cowboy

¹Reproduced with permission. The stimulus pictures were created by Albena Kapte bileva, whose moral rights are hereby acknowledged. For further information about image copyright and conditions for reuse of the material, please contact the University of Oldenburg, http://www.aulin.uni-oldenburg.de/29498.html.
B.1. Exp. 2.1: Active/passive sentence priming

3. a) Waiter interviewing hairdresser  
   b) Hairdresser interviewing waiter

4. a) Tourist interviewing knight  
   b) Knight interviewing tourist

5. a) Robber chasing magician  
   b) Robber chasing magician
6. a) Tourist chasing painter  
   b) Painter chasing tourist

7. a) Policeman shooting chef  
   b) Chef shooting policeman

8. a) Ghost filming magician  
   b) Magician filming ghost
B.1. Exp. 2.1: Active/passive sentence priming

9. a) Tourist interviewing girl  
   b) Girl interviewing Tourist

10. a) Nun chasing robber  
     b) Robber chasing nun

11. a) Pope touching detectives  
      b) Detectives touching pope
B. Stimulus material study 2

12. a) Vampires observing painter
    b) Painter observing vampires

13. a) Chef shooting tourists
    b) Tourists shooting chef

14. a) Baker interviewing robbers
    b) Robbers interviewing baker
B.1. Exp. 2.1: Active/passive sentence priming

15. a) Monster filming hairdressers  
   b) Hairdressers filming monster

16. a) Grandmother filming queen  
   b) Queen filming grandmother

17. a) Doctor greeting carpenters  
   b) Carpenters greeting doctor
B. Stimulus material study 2

18. a) Farmers greeting doctor b) Doctor greeting farmers

19. a) Soldiers reprimanding tourist b) Tourist reprimanding soldiers

20. a) Waitress shooing away fairies b) Fairies shooing away waitress
B.2. Exp. 2.2: Relative clause elicitation

Stimulus sentences

Each picture pair was combined with stimulus sentences containing subject relative clauses (SR) and object relative clauses (OR).

1. **SR:** Hier lächelt der Kapitän, der die Diebe berührt.
   “Here the captain, who is touching the thieves, is smiling.”
   **OR:** Hier lächelt der Kapitän, den die Diebe berühren.
   “Here the captain, whom the thieves are touching, is smiling.”

2. **SR:** Hier lächelt der Mönch, der die Astronauten erschießt.
   “Here the monk, who is shooting the astronauts, is smiling.”
   **OR:** Hier lächelt der Mönch, den die Astronauten erschießen.
   “Here the monk, whom the astronauts are shooting, is smiling.”

3. **SR:** Hier lächelt der Friseur, der die Köchinnen erschießt.
   “Here the hairdresser, who is shooting the female chefs, is smiling.”
   **OR:** Hier lächelt der Friseur, den die Köchinnen erschießen.
   “Here the hairdresser, whom the female chefs are shooting, is smiling.”

4. **SR:** Hier niest der Reporter, der die Vampire filmt.
   “Here the journalist, who is filming the vampires, is sneezing.”
   **OR:** Hier niest der Reporter, den die Vampire filmen.
   “Here the journalist, whom the vampires are filming, is sneezing.”

5. **SR:** Hier schwitzt der Zimmermann, der die Piloten grüßt.
   “Here the carpenter, who is greeting the pilots, is sweating.”
   **OR:** Hier schwitzt der Zimmermann, den die Piloten grüßen.
   “Here the carpenter, whom the pilots are greeting, is sweating.”

6. **SR:** Hier zittert der Taucher, der die Zauberer malt.
   “Here the diver, who is painting the magicians, is trembling.”
   **OR:** Hier zittert der Taucher, den die Zauberer malen.
   “Here the diver, whom the magicians are painting, is trembling.”

7. **SR:** Hier weint der Kellner, der die Matrosen tadelt.
   “Here the waiter, who is reprimanding the sailors, is crying.”
   **OR:** Hier weint der Kellner, den die Matrosen tadeln.
   “Here the waiter, whom the sailors are reprimanding, is crying.”

8. **SR:** Hier lächelt der Bauer, der die Ärztinnen fängt.
   “Here the farmer, who is catching the female doctors, is smiling.”
   **OR:** Hier lächelt der Bauer, den die Ärztinnen fangen.
   “Here the farmer, whom the female doctors are catching, is smiling.”
B. Stimulus material study 2

9. SR: Hier lächelt der Kapitän, der die Männer filmt.
   “Here the captain, who is filming the men, is smiling.”
   OR: Hier lächelt der Kapitän, den die Männer filmen.
   “Here the captain, whom the men are filming, is smiling.”

10. SR: Hier niesen die Königinnen, die die Hexe erschießen.
    “Here the queens, who are shooting the witch, are sneezing.”
    OR: Hier niesen die Königinnen, die die Hexe erschießt.
    “Here the queens, who the witch is shooting, are sneezing.”

11. SR: Hier schwitzen die Königinnen, die die Nonne fangen.
    “Here the queens, who are catching the nuns, are sweating.”
    OR: Hier schwitzen die Königinnen, die die Nonne fängt.
    “Here the queens, who the nun is catching, are sweating.”

12. SR: Hier weinen die Cowboys, die die Hexe interviewen.
    “Here the cowboys, who are interviewing the witch, are crying.”
    OR: Hier weinen die Cowboys, die die Hexe interviewt.
    “Here the cowboys, who the witch is interviewing, are crying.”

13. SR: Hier schwitzen die Kapitäne, die die Ärztin verfolgen.
    “Here the captains, who are chasing the female doctor, are sweating.”
    OR: Hier schwitzen die Kapitäne, die die Ärztin verfolgt.
    “Here the captains, who the female doctor is chasing, are sweating.”

14. SR: Hier gähnen die Bräute, die die Prinzessin treten.
    “Here the brides, who are kicking the princess, are yawning.”
    OR: Hier gähnen die Bräute, die die Prinzessin tritt.
    “Here the brides, who the princess is kicking, are yawning.”

15. SR: Hier erröten die Bäuerinnen, die die Fee wecken.
    “Here the female farmers, who are waking the fairy, are blushing.”
    OR: Hier erröten die Bäuerinnen, die die Fee weckt.
    “Here the female farmers, who the fairy is waking, are blushing.”

16. SR: Hier zittern die Mönche, die das Model bestehlen.
    “Here the monks, who are stealing from the model, are trembling.”
    OR: Hier zittern die Mönche, die das Model bestiehlt.
    “Here the monkey, who the model is stealing from, are trembling.”

17. SR: Hier niesen die Köche, die die Touristin erschießen.
    “Here the cooks, who are shooting the female tourist are sneezing.”
    OR: Hier niesen die Köche, die die Touristin erschießt.
    “Here the cooks, who the female tourist is shooting, are sneezing.”
18. *SR:* Hier lächelt die Putzfrau, die die Gärtnerinnen fängt.
   “Here the cleaning lady, who is catching the female gardeners, is smiling.”
   *OR:* Hier lächelt die Putzfrau, die die Gärtnerinnen fangen.
   “Here the cleaning lady, whom the female gardeners are catching, is smiling.”

19. *SR:* Hier lächeln die Köche, die die Nonne interviewen.
   “Here the cooks, who are interviewing the nun, are smiling.”
   *OR:* Hier lächeln die Köche, die die Nonne interviewt.
   “Here the cooks, who the nun is interviewing, are smiling.”

20. *SR:* Hier weinen die Räuber, die die Lehrerin tadeln.
   “Here the robbers, who are reprimanding the female teacher, are crying.”
   *OR:* Hier weinen die Räuber, die die Lehrerin tadelt.
   “Here the robbers, who the female teacher is reprimanding, are crying.”
B. Stimulus material study 2

Stimulus pictures

1. a) Smiling captain touching thieves  
   b) Thieves touching smiling captain

2. a) Smiling monk shooting astronauts  
   b) Astronauts shooting smiling monk

²Reproduced with permission. The stimulus pictures were created by Albena Kaptebileva, whose moral rights are hereby acknowledged. For further information about image copyright and conditions for reuse of the material, please contact the University of Oldenburg, http://www.aulin.uni-oldenburg.de/29498.html.
B.2. Exp. 2.2: Relative clause elicitation

3. a) Smiling hairdresser shooting chefs
   b) Chefs shooting smiling hairdresser

4. a) Sneezing journalist filming vampires
   b) Vampires filming sneezing journalist

5. a) Sweating carpenter greeting pilots
   b) Pilots greeting sweating carpenter
B. Stimulus material study 2

6. a) Trembling diver painting magicians
   b) Magicians painting trembling diver

7. a) Crying waiter reprimanding sailors
   b) Sailors reprimanding crying waiter

8. a) Smiling farmer catching doctors
   b) Doctors catching smiling farmer
9. a) Smiling captain catching men  
   b) Men catching smiling captain

10. a) Sneezing queens shooting witch  
    b) Witch shooting sneezing queens

11. a) Sweating queens catching nun  
    b) Nun catching sweating queens
B. Stimulus material study 2

12. a) Crying cowboys interviewing witch
    b) Witch interviewing crying cowboys

13. a) Sweating captains chasing doctor
    b) Doctor chasing sweating captains

14. a) Yawning brides kicking princess
    b) Princess kicking yawning brides
15. a) Blushing farmers waking fairy  b) Fairy waking blushing farmers

16. a) Trembling monks stealing from model  b) Model stealing from trembling monks

17. a) Sneezing cooks shooting tourist  b) Tourist shooting sneezing cooks
B. Stimulus material study 2

18. a) Smiling cleaner catching gardeners
   b) Gardeners catching smiling cleaner

19. a) Smiling cooks interviewing nun
   b) Nun interviewing smiling cooks

20. a) Crying robbers reprimanding teacher
   b) Teacher reprimanding crying robbers
C. Stimulus material study 3

   “The complaint/s by the dedicated student/s is/are justified.”

   “The figure/s in the colorful brochure/s is/are clear.”

   “The inscription/s on the ancient pillar/s is/are weathered.”

   “The round/s of the surgical ward/s is/are finished.”

5. Die Affäre/n um den/die berühmten Spion/e ist/sind übertrieben.
   “The scandal/s about the famous spy/spies is/are exaggerated.”

   “The picture/s on the sought-after cup/s is/are pretty.”

   “The newspaper/s with the embarrassing mistake/s is/are untraceable.”

   “The gap/s in the awkward theory/theories is/are highly visible.”

   “The cooperation/s with the commercial dealer/s is/are fruitful.”

    “The traditional costume/s with the typical pattern/s is/are sought-after.”

    “The description/s in the up-to-date travel guide/s is/are unclear.”

    “The direction/s in the internal newsletter/s is/are inaccurate.”
C. Stimulus material study 3

   “The manipulation/s of the complex machine/s is/are dangerous.”

   “The display/s in the renovated train/s is/are lit.”

   “The gathering/s in front of the closed store/s is/are unexpected.”

   “The signature/s on the binding declaration/s is/are forged.”

   “The treatment/s with the novel drug/s is/are effective.”

   “The accident/s in the processing factory/factories is/are catastrophic.”

   “The change/s to the effective regulation/s is/are minimal.”

    “The demand/s for the substantial reform/s is/are unsound.”
Bibliography


Bibliography


Bibliography


Bibliography


Delage, H. & Tuller, L. (2010). Evolution of syntactic complexity and avoidance strategies in children and adolescents with mild-to-moderate...
Bibliography


agreement. *Journal of Memory and Language*, 36, 147–164.


Bibliography


Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C.,


Garrett, M. (1980). Levels of processing in sentence production. In B. Butter-
Bibliography


Bibliography


Heuer, H. (1996). Doppeltätigkeiten [Dual tasks]. In O. Neu-


Bibliography


Bibliography

Dordrecht: Kluwer.


Bibliography


Bibliography


Bibliography


Bibliography

*Language, 37, 440–479.*
Schütze, C. T. & Gibson, E. (1999). Argumenthood and English prepos-


Bibliography


listeners. *International Journal of Audiology, 45*(1), 26–33.


Zusammenfassung


Gerade die Generierung von syntaktischer Struktur während der Sprachproduktion wurde in der Vergangenheit als ein Beispiel für hochgradig automatische kognitive Verarbeitungsprozesse angesehen. Im Hinblick auf das intuitive Empfinden einer höheren ‘Anstrengung’ beim Sprechen in akustisch schwierigen Situationen stellt sich die Frage, ob dem Sprechen tatsächlich so isolierte kognitive Prozesse vorausgehen, wie es in der Vergangenheit häufig angenommen wurde.

Zusammenfassung

die theoretischen Grundlagen eines Satzproduktionsmodells diskutiert. Die theoretische Diskussion schließt an das vorhergehende Kapitel an und behandelt unter anderem die Frage nach dem Verhältnis zwischen linguistischer Theorie und Verarbeitungsmodellen aus einer psycholinguistischen Perspektive.

Ausgehend davon wird diskutiert, wie der Begriff der linguistischen Komplexität substantiiert werden könnte, und in welcher Beziehung ein theoretiegebundener Komplexitätsbegriff zu Modellen eines Verarbeitungsmechanismus steht. Es wird die Schlussfolgerung gezogen, dass theoretische Konzepte der Komplexität per se keine (eigenständige) Erklärung für Verarbeitungsschwierigkeiten liefern können, da formale Komplexitätsmetriken implizit auf kognitive Verarbeitungsmechanismen rekurrieren, die ein formales Konstrukt für Komplexität überhaupt erst bedingen.

tierungsparadigma untersucht. Die Ergebnisse zeigen eine klare Erhöhung der Fehlerrate unter Störschall.


Erklärung gemäß § 9 Abs. 2 Buchst. b) der Promotionsordnung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

München, 17. Dezember 2013

gez. Mirko Hanke
Lebenslauf Mirko Hanke

Persönliche Daten

geboren am 3. November 1979 in Hanau
Staatsangehörigkeit: deutsch

Ausbildung

2012–2014 Bibliotheksreferendariat an der Staatsbibliothek zu Berlin–Preußischer Kulturbesitz und an der Bibliotheksakademie Bayern, Abschluss mit Staatsexamen

2008–2014 Promotion bei Prof. Dr. Cornelia Hamann an der Carl von Ossietzky-Universität Oldenburg


2003–2004 Auslandsaufenthalt an der Rijksuniversiteit Groningen, NL

2002 Zivildienst am Universitätsklinikum Leipzig

1999–2001 Studium der Theaterwissenschaft und Anglistik an der Universität Leipzig