Sentences in Sentences
Modeling Frequency Effects in Local Syntactic Coherence Processing

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät
der Albert-Ludwigs-Universität Freiburg i. Br.

vorgelegt von

Daniel Georg Müller-Feldmeth
aus Erlangen

Sommersemester 2013

Dekan:
Prof. Dr. Albert Gollhofer

GutachterInnen:
Prof. Dr. Lars Konieczny
Prof. Dr. Evelyn Ferstl
Prof. Dr. Heike Behrens

Datum des Promotionsbeschlusses:
10. Februar 2014

Acknowledgments

This work would not have been written without the help, support and patience of a lot of people. First of all, I want to thank Lars Konieczny whose teaching inspired me to dive into psycholinguistics and connectionism in the first place, and whose linguistic, experimental and statistical expertise has been a substantial support throughout this dissertation. I would also like to thank my dissertation committee of Lars Konieczny, Heike Behrens and Evelyn Ferstl. Special thanks go to Stefan Pfänder who provided absolutely indispensable support during the final stage of writing. I am grateful to Whitney Tabor for many critical and fruitful comments on my work, and for lively and inspiring discussions about the world and other dynamical systems while walking the forests of Germany and Connecticut.

Many aspects of this project emerged from ideas born in the Psycholinguistic Research Lab at the Center for Cognitive Science in Freiburg. The discussions, programming sessions and coffee breaks with Peter Baumann, Sascha Wolfer, Helmut Weldle, Wibke Hachmann, Sarah Schwarzkopf, Anne Karina Feldmeth, and others have been essential building blocks of my work. I want to thank my colleagues from the research training group GRK 1624/1 *Frequency Effects in Language*, in particular my officemate Péter Rácz, as well as the organizers and supervisors, for providing an inspiring research atmosphere and excellent possibilities to develop, rethink and finetune the thesis.

My thanks also go to my parents for their patience and long-lasting support, to my longtime friend Benno Hinkes for being my longtime friend, and to my wife Anne Karina Feldmeth for friendship and patience, as well as technical and emotional support along the way. Finally, I have to thank all the people that contributed to this work and have not been mentioned here.

The dissertation was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), research training group GRK 1624/1 *Frequency Effects in Language*.

Contents

1 Introduction

1.1 Sentences in Sentences: Local Syntactic Coherence as non-existent Garden-Path

(1) Es ist schon spät, als der Student frustriert den Professor anruft.
It is already late, when the student frustratedly the professor calls.
"It is already late when the student frustratedly calls the professor."

On first view, there seems to be nothing special about this sentence and you will not encounter any major difficulties in reading and understanding it: a student calls his professor late in the evening and, for whatever reason, the student is frustrated. However, if we skip the first five words and the main verb, we find the following sequence of words:

(2) [. . .] der Student frustriert den Professor [. . .]
[. . .] the student frustrates the professor [. . .]

If we consider this sequence on its own, it is not the student being frustrated, but the student frustrating the professor (who is now the one being frustrated). As you may have noticed, this only works because «frustriert» is morphologically ambiguous in German: it can occur as an adverbial participle [frustratedly] or as a finite verb in 3rd Person singular [frustrates]. The reason that you probably did not even notice the meaning of this embedded sequence while reading example (1) (even as a native speaker of German) is that «frustriert» can simply not be read as a finite verb *in the context of the whole sentence*. The fragment:

"Es ist schon spät, als der Student frustriert . . . "

does not allow for any meaningful continuation where the student frustrates somebody.

If we, in contrast, read or hear a sentence fragment like:

(3) Der Professor, den der Student regelmässig frustriert [. . .]
The professor that the student regularly frustrates [. . .]

we will probably assume that the student is frustrating for the professor, and will be at least somewhat surprised if the sentence continues with

(4) Der Professor, den der Student regelmässig frustriert anruft, hat die Nase voll.
The professor, that the student regularly frustratedly calls, has the nose full.
"The professor that the student regularly calls frustratedly is fed up."

The latter sentence is *temporally ambiguous*[1] and is an instance of a so-called *garden path* sentence (Bever, 1970): sentences where readers or hearers first consider an interpretation that turns out to be incorrect later – they are "led down the garden path". Garden path sentences are among the most investigated phenomena in psycholinguistics, because they provide insights into the way humans cope with one of the major problems of language: human language is fraught with ambiguity. In severe garden path sentences, as in the classical example

(5) The horse raced past the barn fell.

readers consciously stumble as soon as they read "fell". So, it is evident that ambiguity can sometimes pose problems for the processing system and can even result in judging the sentence as ungrammatical. However, these cases are relatively rare in everyday language, and in the majority of cases we are rather good at coping with ambiguities or do not even notice them at all. Our sentence processing system obviously entails very efficient mechanisms to disambiguate words and sentences and allows us to cope with ambiguity much better than many machine-based language processing systems to date. However, even in cases where comprehenders do not consciously notice any difficulty, we can still measure differences in reading times between sentences with and without local ambiguities.

The fact that we do encounter difficulties or at least react somehow to temporally ambiguous sentences at all, points to a fundamental property of human language processing: If we waited until the end of a sentence and then built up a syntactic analysis and interpretation given all the available information, ambiguity should not be a problem at all. However, this is not the way the human sentence processing system works. In contrast, sentence processing works in an incremental fashion: While processing a sentence, each word we encounter is immediately interpreted and integrated with the words read or heard before, so that we have a partial, but integrated and consistent analysis at every point in the sentence.

The insight – or assumption – that the human parser proceeds *incrementally* emerged not least from psycholinguistic research on ambiguity resolution: We don't wait until the end of a

[1] *Temporal* ambiguity is also often referred to as *local* ambiguity. I will use the term *temporal* mainly to avoid confusion with the phenomenon of *local syntactic coherence*, which is not a case of local *ambiguity*.

sentence to come up with an interpretation, but start to build up an analysis and an interpretation word by word. The classical demonstration of immediate and incremental processing has been provided by Marslen-Wilson (1973), and a number of subsequent studies provided evidence that incremental processing is a fundamental property of human language processing. As a consequence, the majority of current models of sentence processing is based on this assumption.

A lot of details have been under debate, concerning the way this incremental process can be specified. Does the parser proceed serially and thus always goes for one of the possible analyses in case of a temporal ambiguity, or are all or at least some of the possible interpretations pursued in parallel? What sources of information are used to guide the initial decision? However, basically all theories assume that we only *consistent* analyses are taken into consideration.

A partial analysis at word w_n is consistent if it is compatible with the input received in the actual sentence up to w_n. In garden path sentences, there are (at least) two analyses that are compatible with the input before we reach the disambiguating region. The fact that comprehenders only consider consistent analyses seems to be a very plausible hypothesis, irrespective of any detailed theory of parsing: Given the limited cognitive resources humans have, why should we consider analyses or interpretations of sequences of words that are not consistent with the context they are embedded in? Anything else would be a waste of resources. This basic consideration can be formulated as the principle of *self-consistency* (Tabor, Galantucci, & Richardson, 2004) or *global coherence*: Consider only analyses that are consistent with the input received so far. Garden path effects are compatible with this principle, and the principle also matches our intuition about example (1) at the very beginning of this section. We do hardly notice that the sentence entails word sequence that would have another meaning if it occurred in a different syntactic context. So, according to both our intuition and in accordance with most of the current theories of parsing, a word sequence with an alternative interpretation that does not fit with the global (preceding) context, should not be considered and, in consequence, should not induce any difficulty.

However, there is evidence that this very plausible assumption does not seem to be entirely correct: Tabor et al. (2004) were the first to provide evidence that sentences with “merely local syntactic coherence” can indeed result in processing load, measured as elevated reading times. In a self-paced reading experiment, participants read sentences as

(6) The coach chided {the player tossed a frisbee} by the opposing team. (Tabor et al., 2004)

As example (1), this sentence contains a locally coherent sequence, constituted by the morphologically ambiguous element “tossed”: “the player tossed a frisbee”. In the global context of the sentence, however, “tossed” has to be interpreted as a past participle that introduces a reduced passivized relative clause. The results showed that reading times on the locally coherent sequence tend to be longer than in control conditions where the local coherence was prevented either by using an unambiguous verb (“thrown”) or an unreduced relative clause (the player that was tossed a frisbee). Crucially, as example (1), the sentence does not contain a temporal ambiguity, since the finite verb reading is not compatible with the preceding context. It is thus not a garden path sentence – however, readers seem to be led down this strange garden-path although it doesn’t even branch off from the current path they’re on.

This phenomenon of *local syntactic coherence* provides an interesting phenomenon that calls into question not only our intuitions but also fundamental assumptions widely held to be true in psycholinguistic theorizing.

> The majority of parsing research has assumed that people only construct grammatical representations that are consistent with the rules of grammar. Most research has simply assumed that people only construct grammatical representations, and has attempted to determine which grammatical analysis is adopted or favored at different points in processing. The possibility that people construct ungrammatical representations would require psycholinguists to re-think their models and accounts of data, for example, considering the possibility that apparently ungrammatical analyses can interfere with grammatical ones during initial selection or reanalysis. (Pickering, van Gompel, Traxler, & Gernsbacher, 2006, p. 13)

Although empirical data for the phenomenon is still relatively sparse, the phenomenon has attracted a considerable amount of attention in the field. The interest can mainly be attributed to two reasons. On the one hand, the phenomenon calls in question the self-consistency assumption, and thus the view of the language processing system as a *rational* one. On the other hand, the results have been claimed to provide strong evidence for the dynamical-systems perspective on language processing, thus drawing on a fundamental debate that has been occupied cognitive science ever since.

Regarding the first point, as Bicknell and Levy (2009) point out, the assumption of a self-consistent parser draws on very plausible considerations about cognitive processing in general:

Given the boundedness of processing resources and the challenge of real-time processing, it seems to be very plausible that cognitive processes are *rational*: in cases of uncertainty, use all – and only! – information that is available and valuable to come up with the most probable interpretation of the input. Effects of local coherence, however, seem to provide evidence that at least in some cases, the available and helpful information of the preceding context is ignored, rendering the processing of locally coherent sequences irrational.

Regarding the second point, Tabor et al. (2004) take their results as strong evidence against any account that subscribes to self-consistent parsing – and thus against the majority of current computational models of sentence processing. Instead, they argue that the human language processing mechanism is to be conceived as a dynamical *self-organizing* system. Thus, the phenomenon figures as an important piece of evidence in the overarching debate about how we should think about language processing and cognitive processing in general, and what kinds of models (and metaphors) can capture the nature of the human cognitive system most adequately. Is the human cognitive system best described as a computational device that operates on symbols, manipulated through formal rules? Or is it more adequately understood as a highly parallel and interactive, dynamical system that extracts complex patterns from its environment and develops complex behavior in a self-organized, emergent way? These two perspectives on the cognitive system differ significantly, not only in the metaphors they use to give an idea of how human cognition works, but also in the way they conceptualize learning, representation and processing.

1.2 Questions and Hypotheses

To conclude, the phenomenon of local syntactic coherence addresses fundamental questions regarding human language processing, and provides a suitable test case to differentiate between different models and frameworks. However, evidence is still rather sparse and a lot of open questions await investigation. Thus, the aim of this thesis is to delve deeper into the nature of the phenomenon, using it as an important window to get a glimpse on the mechanisms that enable humans to use language. In a nutshell, I will focus on three questions:

1. Can the empirical results reported so far be taken as evidence for influence of local syntactic coherence at all?

2. If so, do they merely pose processing difficulty, or can they even trigger interpretation of the local sequence?

3. Do the effects provide support for a dynamical system perspective on human cognition, as it is, for example, implemented in connectionist approaches of sentence processing?

To put it simply, I will provide evidence that all three questions can be answered by yes. In particular, I will put forward the hypotheses that (a) effects of local syntactic coherence are best explained as locally induced, but globally unwarranted expectations that (b) result from frequent encounter of the local sequence in contexts where the local interpretation is correct, and (c) that these influences can be best accounted for by experience-based self-organizing models of human language processing.

In the following section, I will give an outline of this thesis, spell out why these questions are particularly interesting, and how I will attempt to evaluate the hypotheses.

1.3 Outline of the Thesis

In chapter 2, I will give a rough overview of central topics in research on human sentence processing that are related to the topic of this work. To set the stage, in section 2.1 we will start off with a rather informal introduction into the main issues that sentence processing poses for scientific inquiry. I will then focus on two areas that provide particularly useful windows into the mechanisms underlying human sentence processing: ambiguity resolution (section 2.1.1) and processing complexity (section 2.1.2). This overview will be mainly organized around two major directions that can be identified in both areas: On the one hand, the *working memory* as a limited resource has been put forward as a crucial element to explain effects of ambiguity and as the main factor that determines processing complexity of sentences. On the other hand, linguistic *experience* and the distributional properties of the linguistic environment have been proposed as the main driving force that guides human language comprehension.

Ambiguity resolution In research on ambiguity resolution, the guiding questions regarding the general mechanisms of sentence processing concern the parallel or serial nature of the processing system, and the question whether we should assume different, encapsulated modules of processing or a highly interactive system taking into account information on several structural levels at the same time.

After early theories explaining preferences in ambiguity resolution by proposing particular perceptual strategies (section 2.1.1.1), two-stage models proposed parsing principles relying on structural properties, motivated by considerations of *memory-efficiency* in a rule-oriented computational framework (sections 2.1.1.2, 2.1.1.3). In contrast, interactive constraint-based theories emphasize the importance of probabilistic features of linguistic structures, and are thus experience-based (section 2.1.1.4). The debate about serial or parallel processing and the modular or interactive architecture of the human sentence processing system is still vivid. With regard to experience, it is rather uncontroversial that experience plays a crucial role – however, the exact interplay between syntactic, hierarchical processes and linguistic experience is still widely debated.

As I will argue in section 2.1.1.5, however, the majority of models cannot easily account for effects of local syntactic coherence, mainly because one crucial assumption is that human parsing is strictly incremental and self-consistent (Stabler, 1994; Frazier & Clifton Jr., 1996; Just & Carpenter, 1980), regardless of the various claims regarding properties of the processing system. Thus, locally coherent, but globally inconsistent structural analyses should not have any influence on processing. However, for the most part this is not considered at all (paragraph 2.1.1.5).

Processing complexity Another major strand of psycholinguistic research deals with processing complexity of *unambiguous* sentences. Since I argue that local syntactic coherence is a source of processing complexity and it does not constitute structural ambiguity in the classical sense, I will discuss some approaches that have been put forward in this domain.

Resource-based models argue that processing complexity is a function of the burden a particular element in a syntactic structure places on working-memory, where this burden is determined, for example, by the distance between dependent elements, leading to effects of *locality* (Gibson, 1998, 2000). In more recent approaches, it is not merely distance, but rather the properties of the elements in between dependent elements that can lead to *retrieval interference* and thus to elevated processing time (Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, Johnson, & Lee, 2006; Lewis & Vasishth, 2005). Experience-based approaches, however, argue that processes of anticipation and prediction, based on acquired knowledge about distributional properties of the language, are the main factors that determine processing complexity (e.g. Levy, 2008a; Konieczny, 2000). In direct opposition to Gibson's (1998, 2000) *Dependency locality theory (DLT)*, Konieczny (2000) and Vasishth and Lewis (2006) provided evidence that processing time does not necessarily *increase* with growing distance, as would be predicted by distance based approaches, but can *decrease* if the additional intervening elements provide information that render the dependent element (the verb) easier to anticipate (anti-locality effects). Konieczny (2000) argues that anti-locality effects provide evidence for connectionist SRN-approaches that basically function on prediction given the preceding context.

A phenomenon that plays a prominent role in debates about sentence complexity is the asymmetry in processing effort between subject- and object-extracted relative clauses that has been reported for a number of different languages. In opposition to resource-based models, MacDonald and Christiansen (2002) put forward a radically experience-based account of this difference and describe a connectionist SRN model to support their claim. The modeling approach has been a major inspiration for this project and its predecessors in our lab. In a detailed analysis of the model, Konieczny, Ruh, and Müller (2009) and Ruh, Klöckner, and Konieczny (2002) revealed, on the one hand, some severe weaknesses of the approach, that will have to be taken into account in my own modeling attempt in chapter 4. On the other hand, the analysis revealed that local, ungrammatical predictions play an important role in the model. This observation was one of the main triggers to consider local syntactic coherence as a potential influencing factor in human sentence processing in the first place.

Usage and prediction The crucial role of language use, experience and prediction on several aspects of language processing is incorporated in a variety of different approaches. In section 2.2, I will describe two attempts to provide a broad coverage, expectation based measure of processing difficulty. On the one hand, *syntactic surprisal* (Hale, 2001; Levy, 2008a) is based on the assumption that difficulty is greater if a particular word is unexpected – surprising – given its preceding context. In contrast, *entropy reduction* (Hale, 2003, 2006) formalizes the intuition that a word is more difficult if it entails more information about what the syntactic structure of the whole sentence will finally be. Both measures have been shown to be significant predictors of processing difficulty, as measured in reading times. However, they explicitly subscribe to the self-consistency principle as a rational strategy, since expectations are derived from the entire context of the input processed so far. Thus, although in line with the experience-based approach adopted here, neither measure incorporates expectations based on local context, and thus cannot account for effects of local syntactic coherence.

A rather different instantiation of experience-based sentence processing will be introduced in section 2.2.3, where I will describe simple recurrent networks (SRN, Elman, 1990) as a cognitive model of sentence processing. SRNs, and connectionist architecture that include SRNs as a central component have been used to model a variety of linguistic phenomena, and have proved to provide a simple, but powerful tool to investigate sequential processing. I will describe the basic properties of this architecture and will also point out some problematic aspects that will have to be considered in my own model later on in section 2.2.3.

Evidence and explanations of local syntactic coherence effects Section 2.3 gives a more detailed definition of local syntactic coherence as a descriptive concept, and introduces a number of different examples of the phenomenon (section 2.3.1), followed by an overview of empirical evidence as well as the various theories attempting to account for the data. Firstly I will address the initial results provided by Tabor et al. (2004) and their *self-organized parsing* model SOPARSE (Tabor & Hutchins, 2004). Konieczny (2005) provided further evidence in an anomaly detection experiment. A number of subsequent approaches, as well as some models that did not explicitly address local syntactic coherence, but are potential candidates for the effects will be discussed.

In particular, one group of approaches argues that the effects do not provide evidence for influence of local syntactic coherence as resulting from *context-dependent* bottom-up processes, but can be accounted for by lexical properties of individual words (Gibson, 2006) or by retrieval interference (Van Dyke, 2007). Another group of approaches conceives the effects as resulting from the interaction of global syntactic analysis and a preprocessing module operating on a shortsighted window of several words (Corley & Crocker, 2000; Bicknell & Levy, 2009; Bicknell, Levy, & Demberg, 2009; Frazier & Fodor, 1978). As has been pointed out above, one of the strong claims of Tabor et al. (2004) has been that the results pose a problem for the self-consistency principle underlying most of the current theories of sentence processing. This assumption is partly grounded on the assumption that the human sentence processing mechanism is to be conceived as implementing a *rational* mechanism with regard to the task of comprehension. Consequently, a number of approaches have been put forward to incorporate the effects without abandoning the rationality assumption. Hale (2011) provides a *rational* symbolic parser incorporating the idea that sentence processing – assigning a structural analysis to a surface string – can be conceived as an informed search which can in some circumstances lead to inspection of globally inconsistent states. A different approach has been provided in Levy (2008b). Taking into consideration noise in real-world communication, a rational parsing mechanism also has to account for the possibility that the already processed input was either erroneuos or misperceived. Levy (2008b) shows that for the stimuli used in Tabor et al. (2004), such a mechanism could predict similar results without assuming local coherence processing in the strict sense. The "good-enough parsing approach" (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007) also draws on considerations about "real world" circumstances of communication and provides a further line of explanation: Effects of local syntactic coherence could be a result of a pathway which assigns thematic roles based on heuristics operating on surface properties, that are sufficient in the majority of communicative situations.

The majority of the discussed approaches merely address local syntactic coherence as a potential source of processing complexity, but remain silent about the question if the local sequence may also be *interpreted*. However, in a number of visual world experiments it has been shown that this indeed seems to be the case (section 2.3.6). As I will argue, however, giving the underlying assumptions of the respective models, most of them are rather incompatible with the semantic processing of ungrammatical sequences, or would at least require additional assumptions, even if they can account for differences in reading times. In contrast, in dynamical system based approaches, assuming the highly interactive character of the processing mechanism, influence of a locally coherent sequence on interpretation and thus action is an inherent property of the system, as it is, for example, exemplified in Kukona and Tabor (2011). Thus, the question whether local syntactic coherence influences interpretation can be seen as a particularly interesting question with regard to the underlying mechanisms. In addition, in such approaches we would expect that effects of local coherence are not an *all or nothing* phenomenon, but result from local expectations that are derived from previous experience with the involved sequences. Thus, considering the effects as arising from globally unwarranted, but locally induced expectations, as I argued above, effects of local syntactic coherence should crucially depend on the distributional properties of the particular sequences within the linguistic environment. In other words, local syntactic coherence is to be conceived as a gradient phenomenon, and, consequently, behavioral results should mirror *how locally coherent* a sequence is.

In section 2.4 I will conclude the theoretical part with the main questions arising from the research so far, and specify the particular hypotheses that will guide the empirical part as well as the cognitive model in the next two chapters. The remainder of the thesis will be dedicated to evaluating these hypotheses, on the one hand, by providing empirical evidence (chapter 3) and, on the other hand, by developing a simple recurrent network model (chapter 4).

1.3.1 Two Visual World Experiments and a Corpus Analysis

In chapter 3, I will first describe the visual world paradigm (section 3.1) and argue that it is particularly suited to investigate effects of local syntactic coherence and, in particular, to evaluate if locally coherent sequences trigger interpretational processes (3.2). After a description of the underlying idea how this can be done, I will revisit two experiments implementing this idea. In experiment 1 (section 3.3), participants listened to German complement sentences with or without a locally coherent sequence, which is realized by including either an ambiguous or an unambiguous word, or by adding an additional adverb which renders the local sequence incoherent. While listening to the sentence, participants viewed scenes depicting, among others, the meaning of the local sequence. The results reveal that participants are more likely to fixate the local scene when sentences include a locally coherent sequence. However, the results are subject to some criticisms. In particular, although unlikely, some effects could be interpreted as rather stemming from a potential garden path than from local syntactic coherence (section 3.3.4).

The second experiment addresses this and some other potential issues. Instead of complement clauses, stimuli included dative relative clauses, which rules out the potential garden path phenomenon. The results replicate the fixation patterns observed in experiment 1, thus providing strong evidence that local syntactic coherence does not only induce processing difficulty, as has been shown before, but that the meaning of the local sequence is processed and, at least temporally, maintained, drawing attention to the depicted local meaning. Some observed differences between the results of experiments 1 and 2 are discussed in section 3.4.4.

Section 3.5 evaluates the hypothesis that processing and interpreting local syntactic coherence is an effect of frequency. To support this claim, I conducted a corpus study, measuring the strength of local coherence within the stimuli sentences of experiment 2. Strength is operationalized as the tendency of the ambiguous words to occur as the locally coherent, but globally ungrammatical finite verb meaning. Analyzing the effect of this *finiteness bias (FB)* revealed that fixation proportions are influenced by *FB* such that higher *FB* increases the time spent on the local scene, but only if the preceding context constitutes local syntactic coherence. This supports the hypothesis that the effects arise from frequent experience with the local sequence in contexts where the local interpretation is the correct one.

In section 3.6 I will discuss the results with regard to the different approaches described in chapter 2.3. Word-based explanations, as Gibson (2006) or Van Dyke (2007) can hardly account for the data. For the majority of other approaches, in particular modular ones, effects of interpretation are a rather problematic assumption, although these arguments remain tentative, since this issue is mostly not explicitly addressed. I will argue that the results are best accounted for by dynamical system approaches, as, for example, proposed in Tabor et al. (2004), Kukona and Tabor (2011), and Konieczny (2005).

1.3.2 A Simple Recurrent Network Model of Local Coherence Effects

In chapter 4, I will propose a simple recurrent network model to support the claim that effects of local syntactic coherence result from local expectations arising as *false alarm* activations of ungrammatical, but locally predicted words. Several networks are trained on sentences generated by a probabilistic context free grammar. Training materials are constructed such that the design of visual world experiment 2 can be implemented, as well as the differences in finiteness bias analyzed in section 3.5.

A major part of the chapter will be dedicated to performance on number agreement in long-distance dependencies as a way of assessing if the networks have learned the grammar in the first place (section 4.2). To do this, I will present a method that, in contrast to commonly used performance measures, allows a more comprehensible evaluation of performance on this task in section 4.2.1. This method is based on the comparison of SRN output with a hypothetical model that learns the language perfectly - except number agreement. In section 4.2.2 I will show that the networks perform reasonably well and can thus be regarded as plausible models. Effects of local coherence will be evaluated in section 4.3. In two SRN experiments I will show that effects of local syntactic coherence, measured as false-alarm activations at relevant sentence positions, mirror the results from the visual world experiments, as well as the influence of *FB* (section 4.3.2.1). Using a slightly different design, SRN experiment 2 (section 4.3.2.2) also reveals the expected pattern, but also shows effects that were not readily expected, but may be evaluated in further behavioral experiments. The SRN model provides a rather simple mechanism that can account for effects of local syntactic coherence as a natural outcome of the general mechanisms of language acquisition and processing. This is achieved without stipulating additional modules or particular strategies and thus supports the hypothesis that a dynamical system perspective on sentence processing is particularly suited to incorporate the empirical results.

In the general discussion (chapter 5) I will address some further questions and potential points of criticism, mainly with regard to the modeling approach. Finally, chapter 6 provides a short summary of the achievements and concludes the thesis.

2 Theoretical Background

2.1 Human Sentence Processing

Language processing – understanding and producing written or spoken sentences – is something we accomplish rather effortlessly and, at least in the majority of cases, quite successfully in everyday life. However, as soon as we try to describe the steps and processes that underlie this easy thing, it seems quite mysterious that we are able to accomplish the task at all. The puzzle starts with the question of how we are capable of learning a language in the first place, and does not end with the question of how we cope with the several sources of vagueness and ambiguity language is fraught with. It is even more mysterious if we consider the limited resources of our cognitive system, the often very limited time available for the complex processes and the several sources of noise that can disturb the signal.

What we call an utterance or a sentence, is, physically spoken, just a continuous sequence of sound-waves or a contrast-pattern on a piece of paper or a screen. A lot of things have to be done to derive a meaning from such a pattern. At first, we have to identify specific patterns as language. We have to divide these patterns into meaningful parts like sentences, phrases and words, identify particular word classes, derive the meaning of these words, identify their particular function in the sentence, figure out who does what to whom, which parts depend on which other parts, and which expressions refer to other expressions. Finally, we have to use our knowledge about the world, the speaker, the topic and the situation, what the speaker knows about me, what the speaker thinks I know about him and so forth. If we are lucky, we come up with some interpretation that allows us to react adequately to the signal, maybe by producing sounds, moving our fingers or doing whatever action we consider as an appropriate reaction to the input. Surprisingly, we are lucky most of the time. We can even cope with major disturbances like talking in a discotheque or understanding people with a variety of accents. Usually we can understand what is meant after having heard only parts of a sentence, and most of the time we can cope with irony, metaphors and analogies that have, in their literal sense, not much to do with what is actually talked about. The challenge of psycholinguistic research is to identify the mechanisms and processes that allow us to accomplish all these different tasks in the face of resource- and time-constraints.

Although most of the time this complex task is done quite fast and successfully, the "ease" of understanding differs. Some sentences are easier to understand than others, where "ease" can either be operationalized as *success* or as *processing speed*. In some cases we consciously stumble at a particular location in a sentence and have to reread or ask for clarification, while in other cases, processing is slowed down, but still successful in the end. Systematic differences in ease of understanding are the main source of evidence psycholinguistics can use to gain insight into the underlying mechanisms and processes of the human language processing system. In the following sections, I will focus on two areas that have been particularly fruitful for this enterprise: ambiguity resolution and processing complexity.

Human language is fraught with ambiguity, and a major part of psycholinguists is concerned with the question of how we cope with that difficulty. How do we decide in cases of ambiguity, which one of the possible paths do we choose? Which cases pose how much difficulty and which cases go virtually unnoticed? Which factors influence the decisions and the difficulty, and how can we explain these influences? Nevertheless, even in absence of ambiguity, sentences and different parts of a sentence differ in their complexity – where complexity refers to some measurable difference in processing effort that can, for example, be accessed by measuring reading times. The question is, then, why are some sentences more difficult than others, and what architectural properties, mechanisms and processes can explain these differences?

2.1.1 Ambiguity Resolution

In addition to the general complexity of the task to derive meaning from a sequence of sound-waves or letters, some specific properties of human languages seem to make the task especially hard: Human language is fraught with ambiguity that can occur on a variety of levels. The same string of letters can refer to different word-classes, as in "the old *man* the boat"; the same sound can refer to different entities ("I scream for ice-cream"); phrases can be ambiguous in their referential relations ("the spy watched the man with binoculars"), and so on. In the following section, I will mainly focus on lexical and structural ambiguity, since these two kinds are most central to the topic at hand.

Lexical ambiguity refers to a word having multiple meanings – in example (1) from page 1, the word «frustriert» can occur as a finite verb in the 3rd person singular [the student frustrates somebody], or as an adverbial participle [the student calls frustratedly ...].

(1) Es ist schon spät, als der Student frustriert den Professor anruft.
It is already late, when the student frustratedly the professor calls.
"It is already late when the student frustratedly calls the professor."

Structural ambiguity, in contrast, occurs if a sentence – or a part of a sentence – allows for multiple parses and interpretations. If a complete sentence has multiple interpretations, we will speak of *global ambiguity*. If, however, a sentence has multiple interpretations only up to a particular position in the sentence, but is disambiguated at a later point, I will speak of *temporal ambiguity*. Example (5) from page 2 (repeated below) is an example of temporal ambiguity, since the disambiguating word "fell" rules out the main clause analysis of the preceding sequence.

(5) The horse raced past the barn fell.

In contrast, in a sentence like

(7) Der Student schlägt den Professor mit dem Buch.
The student hits the professor with the book.

there's no disambiguating region: Even after having read the whole sentence, we can not be sure if the student uses the book as an instrument, or if the professor possesses the book. These examples also show that structural ambiguity is often constituted by lexical ambiguity of a single word («frustriert» in example (1)), but can also appear without lexically ambiguous words, as in example (7).

Since ambiguity is such an ubiquitous phenomenon, it has even been argued that the communicative function of language has to be conceived merely as an epiphenomenon of language, since an efficient tool evolved for communicative purposes would not entail as much ambiguity as human languages obviously do (Chomsky, 2002). Piantadosi, Tily, and Gibson (2012), in contrast, argue on information-theoretical grounds that efficiency is exactly the reason why a communication system will be ambiguous, because it allows the re-use of linguistic units and thus increases ease of processing. They point out, however, that the use of the same units for different purposes is only efficient if context information can be easily used to determine the specific purpose of use. Thus, although ambiguity is an omnipresent phenomenon when describing language, it is, as it seems, not as omnipresent when we actually process language. In the majority of cases, ambiguity does not pose any problem, and in most cases we don't even notice it. The reasons are rather obvious: neither words, nor sentences normally appear in isolation, but in context. Regarding lexical ambiguity, in most cases the syntactic context provides clear cues which meaning or word class is intended. The same holds true for temporally and globally ambiguous sentences, where the larger context usually renders one or the other alternative very unlikely.[1] A big part of the puzzle ambiguity poses is thus not the fact that ambiguity causes problems, but rather that it does *not* in the majority of cases.

In the following section, I will give a short overview of some important and influential models that have been proposed to explain ambiguity resolution, and use this phenomenon to get an idea of how human language processing is accomplished.

So, how can we find out how humans accomplish this task, and do this, at least very often, successfully and seemingly effortlessly?

A number of strands have been investigated to answer the question of how the human sentence processing mechanism copes with ambiguity: Do we, in case of a syntactic ambiguity, pursue possible interpretations in parallel, or do we immediately go for one interpretation which we then have to revise at the point of disambiguation? Do we come up with a complete, unique interpretation at all, or can we leave the analysis underspecified and wait for further evidence for disambiguation? What sources of information are available that allow for disambiguation, which ones do we use to arrive at a particular interpretation, and how do the sources of information interact? Although termed with regard to ambiguity resolution, it has not been tried to answer these questions just for the sake of ambiguity, but rather ambiguity resolution has been taken – and proved – as a particular fruitful area to gain a glimpse on the mechanisms and architectural properties of the human language processing system in general. It's neither possible nor necessary to describe the whole range of theories and attempts to explain ambiguity resolution. The following section will thus only give a very rough overview, and I will introduce the main proposals and ideas that have been put forward in the last decades.

2.1.1.1 Perceptual Strategies in Structure Assignment

One of the first approaches that tried to explain syntactic ambiguity resolution and, in particular, attempted to explain initial decisions, was proposed in Bever (1970). The approach was

[1] A lot of cases of ambiguity are only noticed in situations where disambiguating context is lacking: Newspaper headlines, for example, are an excellent source of ambiguous sentences where we sometimes consciously stumble. Reading the headline first, we don't have a very specific context that can help us to disambiguate easily.

in strong opposition to earlier approaches such as Derivational theory of complexity (Miller, 1962), which attempted to prove a close link between processing difficulty and linguistic structures posited within Chomsky's transformational grammar (Chomsky, 1957).

In contrast, Bever (1970) attempted to link general cognitive abilities to linguistic behavior. He proposed several *perceptual strategies*, partly derived from visual processing, to explain the process of language perception as assigning internal sentence structure to surface strings. As an example, one of several proposed perceptual strategies is to initially map *Noun-Verb-Noun* sequences to *Agent-Action-Patient* semantics: This strategy aimed to explain, among others, the severe difficulties comprehenders exhibit when processing sentences like (5). where the strategy induces an initial interpretation that conceives the second noun ("barn") as the patient of "the horse raced".

Bever's approach, however, encounters the problem that a rather large number of strategies have to be assumed to be able to account for the various cases of ambiguity (and other processing differences). Since there's no implemented model incorporating the theory, nor an exhaustive list of strategies that could account for the variety of phenomena, it is particularly hard to evaluate. However, Bever's attempt to link general cognitive abilities to language processing, and to consider the role of rather simple heuristic strategies in assigning linguistic structure to a surface string, has been very influential.

Bevers's (1970) approach also led to a focus on *initial preferences* in parsing, and thus to a concentration on the several forms of ambiguity and their effects on sentence processing.

2.1.1.2 Incorporating Memory Limitations: the "Sausage Machine"

In 1978, Frazier and Fodor published a two-stage parsing model known as the *sausage-machine*. It employs two separate mechanisms: the *Preliminary Phrase packager* (*PPP*), operating on groups of words within the lexical string, that assigns lexical and phrasal nodes to these substrings. This 'shortsighted' mechanism generates structures that are then handed over to the second stage, called the *Sentence structure supervisor* (*SSS*). In this second step the phrasal structures are combined into larger clausal structures, and the *SSS* keeps track of dependencies that are beyond the window of the *PPP*. The underlying motivation for assuming a two-stage mechanism and a limited window for the first stage is the limited capacity of working-memory, prominently posited in Miller (1956). Since it had been shown that more material can be stored in working memory if the items are "packaged", that is, combined into (possibly meaningful) structured *chunks*, the first-stage prepackaging reduces memory load considerably and thus enables the parser to cope with long sentences that would otherwise exceed working-memory capacity. In other words, the sausage machine reduces the number of elements to be considered in the parsing process so that the *SSS* is finally able to represent the complete structure of an otherwise too long sentence. However, since capacity is limited, the parser employs strategies that hold memory load low. If this low-cost variant turns out to be wrong, a sentence will have to be re-parsed and will take longer to be read. A sentence where the low-cost variant is correct will accordingly be easier:

> The parser chooses to do whatever costs it the least effort; if this choice turns out to have been correct, the sentence will be relatively simple to parse, but if it should turn out to have been wrong, the sentence will need to be reparsed to arrive at the correct analysis. (Frazier & Fodor, 1978, pp. 294)

Crucially, attachment preferences are explained as effects of *locality*. Since the *PPP* is "blind" to any items that are beyond (and therefore also before) its window, the parser prefers attachments to constituents that are within that window, which results in the principle of *local attachment*. The sausage model can thus account for several effects of constituent length which were hard to explain for preceding models.

The general strategy of the sausage machine approach is not to simply describe possible principles or strategies that guide human parsing, but rather to explain what can be described as such a strategy as the inevitable consequence of the architecture of the parser, while this architecture in turn is determined by assumptions about general cognitive resource limitations.

The main achievement of the sausage machine can be seen in the assumption that the parser is designed to make best use of the available resources implementing strategies to keep memory load low from the very beginning. In contrast to an earlier model of Kimball (1973), processing complexity in the sausage machine is thus not assumed to result from overload of a limited capacity working memory, but as a side-effect of strategies to avoid this overload. Another important property of the sausage machine is that the first step is explicitly construed as operating on a specific word based string length (the window size), regardless of its syntactic status. However, as the authors admit, the particular size of this window and the grain size this size is defined on has not been readily specified "[The] proper definition [of the capacity of the PPP] is a very interesting question, but we have not attempted to disentangle all of these alternatives" (Frazier & Fodor, 1978, p. 293, footnote 1). The problem of specifying the window size in a way that accounts for the various differences of phenomena turned out to be one of the main weaknesses of the approach. Consequently, the fixed window module was abandoned in later approaches. The general idea of a parser architecture that works incrementally and implements memory-effective strategies that can sometimes lead to processing difficulties or even breakdown, however, has been maintained as the leading idea of subsequent models.

2.1.1.3 Serial and Modular: the Garden-Path Theory

One of the most influential theories that attempted to address the question of initial parsing decisions has been proposed as the *Garden-Path* theory (Frazier, 1987). Frazier's approach also relies on the assumption that grammatical deep structure plays a crucial role in processing, and interacts with the general cognitive architecture, namely, working memory. Frazier formulates two principles that are assumed to guide initial parsing decisions, *minimal attachment* and *late closure* (Frazier, 1987, p. 562):

- Minimal attachment: Do not postulate any potentially unnecessary nodes.
- Late Closure: If grammatically permissible, attach new items into the clause or phrase currently being processed (i.e. the phrase or clause postulated most recently).

Minimal attachment applies to cases where two alternative analysis differ in *structural complexity*: if this is the case, the parser goes for the structurally simpler alternative. Consider a sentence like:

(8) The spy saw the cop with binoculars but the cop didn't see him.

Minimal attachment predicts that the phrase "with binoculars" will be initially attached to the VP "saw", resulting in a flat [NP [VP NP PP] ...] structure. The alternative (and finally more plausible) parse, however, requires an additional node: [NP [VP [NP PP]] ...] and is therefore not considered at first. The principle of minimal attachment, however, is based on differences in structural complexity and can not explain initial decisions in cases where both possible structures require the same number of nodes. Consider the sentence:

(9) He will read the paper that he received tomorrow.

Both the initially preferred reading, that then turns out to be wrong (He received the paper tomorrow: [NP [VP NP [RC NP VP PP]]]) and the finally correct one (He will read the paper tomorrow: [NP [VP NP [RC NP VP] PP]]) contain the same number of nodes. The principle of late closure posits that the parser tends to keep the currently processed phrase (VP "received") open as long as possible, with the consequence that incoming material will be attached to the most recent phrase. In our example (9), this results in a first attempt to attach "tomorrow" to "received". Both examples are locally ambiguous, and the initial decision turns out be wrong or at least not the most plausible one. As soon as the disambiguating region is reached, it is argued, the parser has to revise its initial analysis, resulting in increasing processing cost in comparison to sentences that do not require reanalysis.

Some crucial assumptions underlying the garden-path theory have to be pointed out. The principles are based on a theory of parsing that assumes a *serial* and *modular* architecture of the human language processor, at least concerning the parser. If more than one possible analysis is compatible with the input, the parser initially goes for only one of the possible paths, and does not maintain alternative parses at the same time. The consequence of this serial process is that if this initial decision proves to be wrong at a later stage, reanalysis has to occur to finally arrive at the correct interpretation. Modularity is incorporated in the assumption that only a very limited level of information – namely, structural information – is used to guide the initial analysis: the principles are solely formulated with respect to the syntactic structural properties of the input. Other sources of information (e.g., lexical and semantic information) do not contribute to the initial decision, but come in only at a later processing stage. This assumption of *syntax first* assumes encapsulated modules, and implements a two-stage model, where the first stage refers to syntactic analysis, while the second stage (the thematic processor) uses semantic and other information if the initial analysis does not succeed.

These assumptions are grounded on considerations about the nature of human memory on the one hand, and general assumptions about linguistic structure on the other. The two-stage model of parsing reflects the idea that a system with limited working memory will tend to initially do as little work as possible to arrive at a first guess.

In contrast to both Bever's (1970) approach and the sausage machine, the Garden-path theory implies strictly incremental processing: No step in the parsing process incorporates operations that are "blind" to the global structure that has been built up starting from the beginning of the sentence – at least regarding structural information. The approaches, however, share the assumption that initial parsing decisions are guided by a set of general principles that are based on structural properties and that lead to a consistent preference for one interpretation over other possible ones. Other sources of information – e.g. plausibility – only come in at a later stage in processing.

These basic assumptions have been called in question by interactive-constraint based theories, that provide a fundamentally different approach to explain ambiguity resolution and will be discussed in the next section.

2.1.1.4 Parallel and Interactive: Interactive Constraint-Based Approaches (ICT)

The basic idea of interactive constraint-based parsing is that a variety of constraints are integrated to activate the most likely interpretation of the input string. Crucially, these constraints are typically probabilistic in the sense that they incorporate comprehenders' *experience* with language and its distributional properties. At any point in the sentence, the probabilistic constraints provide cues as to how likely a particular analysis of the input is. Parallel processing is incorporated by the fact that, in contrast to serial models, not only is one analysis considered at a time, but rather all of the possible ones. The one that is satisfies the various constraints best, will finally be selected. Interactivity is implemented by the fact that all constraints operate at the same time, and that the constraints draw on a wide variety of levels, including, depending on the particular model, frequency of occurrence on different levels of granularity, frequency of particular constructions, frequency of particular lexical items in a construction and so forth.[2]

Interactive constraint-based approaches emerged due to findings that directly questioned the assumptions of seriality and modularity of the human parser. In particular, several studies (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994) showed that semantic and lexical properties influence preferences in ambiguous sentences very early, thus questioning the assumption that initial decisions are mainly or exclusively based on structural, syntactic properties of the input string, as proposed by modular syntax first models.

The rise of interactive constraint-based models is intimately linked with the connectionist approach that regained attention in the late 1980s and early 1990s (For a review of the development of ICT approaches and a review of evidence supporting it see MacDonald, Seidenberg, Traxler, & Gernsbacher, 2006). Interactive constraint-based models of sentence processing thus question both seriality and modularity of the human parser by positing that the human sentence processing system is:

1. *parallel*: more than one possible interpretation/structure can be active at the same time

[2]It is important, however, that the following questions (a) whether the parser employs a modular or an interactive architecture, (b) whether it proceeds serial or parallel and (c) what particular sources of information are used for parsing decisions are independent of each other, although they often come in typical combinations (see Gibson & Pearlmutter, 2000, for a discussion).

2. and *interactive*: a broad range of sources are used to guide parsing decisions, ranging, depending on the model, from lexical information, co-occurrence of words/phrases, plausibility, extrasentential context and others.

The crucial property of ICTs is that the strength or weight of the multiple, interacting constraints on parsing decisions is determined by a range of *probabilities* derived from language input.

Apart from the focus on experience as a crucial factor in language processing, the constraint satisfaction mechanism also implies a different explanation of difficulty when confronted with ambiguity. In principle-based serial approaches, the parser picks out only one of the possible parses, and if this parse turns out to be wrong, time-consuming reanalysis has to occur at the disambiguating region to finally arrive at the correct interpretation. If, as is proposed in interactive constraint-based models, alternative parses are activated in parallel, elevated reading time stems from *competition* between alternatives, leading to longer times until one alternative is finally selected.

The consequence is that in case of an ambiguity we should expect higher processing load and thus elevated reading times at the ambiguous region. van Gompel, Pickering, and Traxler (2001) however, showed that in reading data, competition effects do not show up. They also showed that non-structural information influences processing very early on, thus being in conflict with the modularity assumption proposed by modular syntax-first models. Their *unrestricted-race* model allows – in contrast to parallel models – for only one analysis, but assumes that several sources of information influence the decision (see also van Gompel, Pickering, Pearson, & Liversedge, 2005).[3]

Another drawback has been pointed out with regard to the possibility to evaluate – or falsify – interactive-constraint-based approaches: While it is relatively easy to derive and test or falsify predictions from general principles that are restricted to a particular level of linguistic structure, the complex interaction of multiple weighted constraints is much harder to evaluate. As Pickering et al. (2006) point out, constraint-based models tend to be hard to evaluate, as long as individual constraints are not explicitly spelled out. A number of studies thus attempted to identify which factors influence ambiguity resolution, and parsing difficulty in general. MacDonald et al. (2006) provide a list of constraints that have been shown to influence parsing preferences and reading difficulty in a broad range of structures.

[3]See Vosse and Kempen (2009) for a defense of ICT with regard to these claims.

Several constraint-based models propose a competition mechanism based on lexical entries containing rich information on syntactic, semantic and pragmatic levels that guide the dynamics and results of the parsing process. Although similar in the basic assumptions, a more radical approach to implement an interactive, parallel processing system that is based on competition has been proposed by Tabor, Juliano, and Tanenhaus (1997). The authors provide a connectionist model, consisting of a simple recurrent network and an attached attractor-based network. The *Visitation Set Gravitation model* lacks any explicit notion of “lexical entry” or working memory in the classical sense. Reading times are modeled as the cycles that it takes the attractor-network to settle into a stable state, leading to longer times if different meanings of a sentence put the system in a state that is more distant from the attractor representing the correct parse.

2.1.1.5 Preliminary Discussion

Seriality and modularity This outline only touched upon some of the manifold issues discussed in the field of ambiguity research. In particular, both the questions of modularity and the issue of the serial or parallel nature of the human sentence processing mechanisms have been discussed extensively, and have not been definitively answered. However, it is rather uncontroversial that sentence processing in general proceeds incrementally, integrating incoming material immediately while processing the input. Secondly, there is rich evidence that a lot of different sources of information on different linguistic levels influence which alternative is selected in case of ambiguity, and that a variety of different sources come into play very early during processing.

Experience and working memory The crucial role of distributional properties has been incorporated in virtually all recent models of sentence processing, regardless of the particular mechanism that is assumed to explain human parsing. However, while in grammar-oriented approaches frequency comes into play as an additional source on top of or interacting with grammatical rules, in interactive-constraint based approaches it is the very foundation of both the processing mechanism and, crucially, of grammar itself. Perhaps most radically, the central role of frequency as the very foundation of both language acquisition and processing is inherent in connectionist processing models, as in Tabor et al. (1997), where the model lacks any explicit representation of rules or symbolic representation. Here, sentence processing is not conceived as the task of deriving a symbolic representation of the sentence meaning by successive application of rules on the input elements. Rather, “comprehension” is construed as a successive change in a complex system of interdependent continuous variables, finally resulting in a system state that allows the system to react adequately.

The different approaches are, as we have seen, reflected in the arsenal of concepts that are put forward as explanatory elements of the theories. An approach that assumes discrete symbolic representations requires some form of “container” where the representations are stored, and where intermediate results of processing are held available for further manipulation. This account of memory and, in particular, working memory functions as the main motivation for memory efficient principles and as the bottleneck that is put forward as an explanation of processing difficulty and parsing preferences.

The principles formulated in models like Kimball (1973), Frazier and Fodor (1978) and Frazier (1987) are motivated by consideration of memory resources. As Frazier and Fodor (1978) note, the principles mirror the limited capacity of working-memory in the sense that they allow processing of complex sentences in spite of memory limits. The *PPP* in the sausage-machine model is thought of as a solution of the processing of sentences exceeding working memory capacity. Similarly, the principles of late closure and minimal attachment in the Garden-path theory reduce working-memory load. In interactive constraint-based models, however, working memory, conceptualized as the place to temporarily store the elements of the actual sentence, does not play a prominent role. In contrast, these models emphasize the role of frequency, since constraints are weighted by probabilities extracted from the linguistic environment. Thus, the main force that shapes the system and its behavior are not memory limitations, but the adaptation to the linguistic environment. The overall strategy of the system thus is not conceived as “keeping memory load low” but as “take the most probable road”. Of course, in some sense both ways subscribe to the plausible assumption that the human sentence processing mechanism tends to take the easiest route - they differ, however, fundamentally in the way “ease” is conceptualized.

What about local syntactic coherence? Despite the various models and open questions, research on ambiguity resolution has put forward a crucial property of human sentence processing: While reading or listening to a sentence, in general we do not wait until we have enough information to solve potential ambiguities. In contrast, processing is incremental, thus attempts to immediately integrate incoming material in a coherent representation of the partial input. The majority of models also subscribe to a principle that is, however related to incrementality, even more constraining: Parsing not only works incrementally, but this incremental process of interpretation draws on a particular context, which is the current *sentence*. The set of parses from which one finally will be selected (be it by constraint-satisfaction or structure-based principles or heuristics) only contains *self-consistent* parses, analyses that are consistent with the input that has been processed so far, while "so far" refers (at least) to the beginning of the current sentence. This principle is reflected in the very definition of ambiguity: a sentence is only considered as structurally ambiguous if a sentence allows for multiple interpretations *given the input already perceived*. The principle of *self-consistency* is one of the reasons why effects of local syntactic coherence are not predicted or even considered in the majority of models.

However, the perceptual strategies approach (Bever, 1970) could in principle incorporate the misleading interpretation of a locally coherent string: Given the strategy to interpret any *Noun-Verb-Noun* structure as an *agent-action-patient* structure, this strategy would in principle allow it to be applied in a local, "context-ignorant" way. However, in case of a local syntactic coherence constituted by a morphologically ambiguous element, as in example (1), the POS-sequence is not readily available: assigning *Noun-Verb-Noun* to the sequence "der Student frustriert den Professor" somehow has to be accomplished prior to the application of the strategy. If this assignment is based on the likelihood of occurrence of the surface string with a particular POS-sequence, effects of local coherence should emerge. However, since there's no detailed implementation available, this assumption remains rather tentative.

Although phenomena of local syntactic coherence have not been explicitly considered in the sausage machine model, the shortsighted *PPP* would in principle mean that locally coherent substrings pose problems and, in the framework of the approach, would induce later reanalysis. Given that the locally coherent sequence happens to be prepackaged as a whole, the ignorance of the *PPP* regarding the preceding context would inevitably trigger reanalysis as soon as the prepackaged structure can not be integrated with the rest of the syntactic structure by the *SSS*. However, the sausage machine and, in particular, the assumption of the *PPP* as a shortsighted module operating on a limited number of surface elements have been abandoned in favor of a more parsimonious model, not least because a reasonable specification of window-size and grain-size turned out to be problematic.

The Garden-Path model, abandoning the fixed-window *PPP* module, does not predict any effect of local syntactic coherence. The principles are clearly formulated under the assumption that the parser has a consistent structural representation available at each processing step.

The case is a bit difficult regarding interactive constraint-based models. As far as I can tell, in most interactive constraint-based approaches, the principle of global consistency is at least tacitly assumed. Kamide, Altmann, and Haywood (2003), arguing strongly for an interactive approach, write:

> We suggest that such processing is the hallmark of an incremental processor that is able to draw on different sources of information (some non-linguistic) at the earliest possible opportunity to establish the fullest possible interpretation of the input at each moment in time. (Kamide, Altmann, & Haywood, 2003, p. 133)

The claim that different sources are used at the "earliest possible opportunity to establish the fullest possible interpretation of the input at each moment in time" at least implies the assumption that this interpretation does not include partial analyses that ignore some of these informational sources. It does, however, not necessarily exclude the possibility that partially inconsistent parses are maintained in parallel to the "fullest possible" one.

In general, since local syntactic coherence per definition does not fall under cases of temporal or global ambiguity, the majority of theories does not explicitly consider the phenomenon at all. However, any theory that subscribes to the assumption that parsing is self-consistent will not be able to account for the phenomenon. Effects of local syntactic coherence, however, provide evidence that self-consistency may be not as dominant as it is often assumed. There may be cases where other constraints are strong enough to lead to temporal deviation from the optimal path that self-consistency should constitute. A more detailed discussion of the attempts to explain effects of local syntactic coherence on the basis of the different approaches will be given in section 2.3.

2.1.2 Processing Complexity

The concurring paradigms underlying models of ambiguity resolution as well as the particular factors they focus on are, of course, not only at stake in ambiguity resolution. With regard to sentence processing, another area where these paradigms and thus the focus on working-memory and experience has played a prominent role is research on processing of complex, but unambiguous sentences. As a consequence of the different conceptualizations of "ease" that are taken as the general force of the processing system, the different approaches provide fundamentally different explanations on what induces processing difficulty. The following section will give a glimpse into this strand of research.

As we have seen in the preceding section the various types of ambiguity and the their effects on language processing have provided a crucial source of psycholinguistic theorizing and model building and led to important insights. However, resolving ambiguities is not the only task the human sentence processor has to fulfill. Research on ambiguity has been very fruitful to test particular theories and identify factors that influence sentence processing but does not necessarily answer the question of how language processing is possible at all and may not give insights into the mechanisms that underlie sentence processing in the absence of ambiguity. Thus, another main topic is *processing complexity* of sentences – some sentences are more difficult than others.

What makes a sentence difficult? If we try to think of a complex or difficult sentence, it will probably be long rather than short, and it will most likely include some degree of embedding: While "The dog chases the cat" doesn't pose any problems, the sentence "The dog that the man that the boy sees shouts at chases the cat." is hard. But what exactly makes the latter sentence more complex? Is it just due to the fact that we have to remember more things – agents and actions – in the latter? What about the sentence: "The boy sees the man that shouts at the dog that chases the cat." It is still harder than the short one, but we don't have as many difficulties figuring out who does what to whom in this case, compared to the second one, although the sentence carries roughly the same pieces of information, the same number of actions and agents as the second one. The number of things to remember does not differ, but still the second sentence is far more difficult. Thus, the way dependent elements are organized seems to constitute a considerable factor of complexity. However, we can construct a structurally very similar sentence: "The planet that the astronomer that the university hired saw exploded". While having the same syntactic structure, our knowledge about what kind of agents normally do what kind of actions helps us greatly in interpreting (or rather: guessing?) the sentence. On the other hand, this knowledge can also lead us down the wrong path, or at least make it more difficult to get on the right track, as in a sentence like "The policeman that the spy that the thief observed arrested had been stolen the car". Thus, semantic and pragmatic properties obviously interact with structural properties of the sentence.

Differences in processing complexity thus provide another window into the mechanisms and processes underlying sentence comprehension: simply put, if one sentence is more difficult to process than another, there must be a reason that causes this difference. Thus, identifying factors that influence processing complexity can shed light on the mechanisms that give rise to the influence of these factors. However, the most interesting cases to gain insights into the processes and mechanisms involved in sentence processing are not differences in comprehension success, but rather in slight differences in processing difficulty and cognitive effort while the sentence is being processed.

In the following section, I will give a sketch of some approaches to sentence complexity. I will mainly focus on the opposition between memory-based approaches and experience-based approaches. A main topic in this area of research xn the processing of relative clauses and, in particular, differences between subject- and object-extracted relative clauses, where these (and other) approaches have been extensively discussed.

Although the ability to use and comprehend language has often been claimed to be a very specific cognitive ability, it is nevertheless uncontroversial that theories and models of this ability have to be compatible with or embedded in more general cognitive theories and the architecture and properties of the cognitive system of humans in general.

As pointed out in section 2.1.1, two basic properties of human cognition have been put forward as being major determinants of the nature of human sentence processing. On the one hand, in computational approaches *working memory* plays a crucial role in cognitive processing in general, and in language processing in particular. Language processors have to cope with the sequential structure of language and with the fact that the meaning of a sentence is constituted by the relation between its elements. A sentence unfolds over time – where this unfolding over time is either due to the sequential nature of speech input or, while reading, due to the sequential nature of eye movements, adding the dimension of time to the spatial nature of written text. Therefore, we have to keep certain elements or referents in memory and integrate them to come up with an interpretation. As a consequence, a sentence is likely to be more difficult where a lot of integration has to be done, and if the *distance* between the elements that have to be integrated is large. The differences in working-memory burden that different sentences imply have been the foundation of a wide range of attempts to explain and model sentence complexity. On the other hand, cognitive processing is fundamentally shaped by the *experience* we accumulate during our lifetime. One aspect of experience with regard to language can be construed as acquiring (implicit or explicit) knowledge about the distributional structure of a language. Given the sequential nature of linguistic input, we can, based on our acquired knowledge about the distributional structure of language, develop *expectations* about what will happen next. As a consequence, a sentence or a specific element in the sentence will be difficult if it does not match the predictions or expectations that we derive from our former experience with language. Experience-based accounts thus draw on another fundamental mechanism that is omnipresent in human cognition: We are constantly guessing what will happen next, and these guesses are typically informed both by the actual situational context and, crucially, by probabilistic knowledge about events we have encountered in the past.

Of course, experience is intimately related to memory as well. The two approaches, however, can be understood as focusing on different time scales. Working memory approaches focus on rather short-term processes – interference of elements that have to be retrieved from memory, decay of activated elements over time, or the amount of information that can be kept in working memory. Experience, on the other hand, focuses on long-term processes that build on knowledge acquired long before the actual utterance. Although it is uncontroversial that both time scales play a role, the perspectives often provide different explanations for phenomena of sentence complexity, emphasize different factors that influence processing, and place different mechanisms and processes at the core of language processing.

A crucial property of human language is that different parts in a sentence depend on each other, and to get these dependencies right is a fundamental condition to be able to understand a sentence. So, for example, it is essential to integrate a verb with its subject and its potential objects to understand who does what to whom. As we all know from reading elaborated literature, juridical texts or newspapers, this isn't always an easy task - even if the dependencies are unambiguous. Similarly to the issue of ambiguity, this problem is most evident in cases where we consciously stumble – when our automatic mechanisms seem to break down. The general task, however, has to be done in every sentence. In contrast to sentences containing ambiguities, the problem is not that there are different possible interpretations that are more or less probable or costly. Thus, the question is not how we come up with the best interpretation, but how we come up with an interpretation at all. The differences in complexity can thus be used as another window into the processes and mechanisms underlying language processing. What makes some sentence more complex – more difficult or time-consuming to process – than others? What can we, based on such differences, conclude about the underlying processes and mechanisms and about the architecture of the human language processing system?

Several empirical studies have shown that in English there's a considerable difference in processing difficulty, mostly measured as elevated reading-times, if we compare subject-extracted relative clauses (SRCs, (10)) and object-extracted (ORCs, (11)) relative clauses (e.g. Gordon et al., 2001; Wanner & Maratsos, 1978; King & Just, 1991).

(10) The reporter who attacked the senator admitted the error.

(11) The reporter who the senator attacked admitted the error.

In subject extracted relative clauses (10) the relative pronoun “that” serves as the *subject* of the RC verb “attacked”, while in ORCs (11) the relative pronoun serves as the *object* of the RC. This contrast provides a rich testbed for theory testing, since, at least in English, the two types of relative clauses can be construed to contain exactly the same lexical elements and thus provide a minimal contrast, only differing in word order. Although the marking of extraction differs among languages, the asymmetry in processing difficulty has been approved in a wide variety of languages and thus is a valuable topic to get a grip on basic mechanisms of sentence processing. A considerable amount of research has been done on processing of relative clauses and, in particular, on the difference between ORCs and SRCs. Since relative clauses are not the main topic of this work, the purpose of the following section is not give a comprehensive review of the different accounts for this particular phenomenon (for recent reviews see Gordon & Lowder, 2012; Staub, 2010). Rather, I will to sketch some of the proposed theories as emphasizing different factors that have been pointed out to influence sentence complexity, and the mechanisms that have been hypothesized as explanations for these influences.

Following Gordon and Lowder (2012), three major approaches to explain the processing asymmetry can be distinguished. Memory/resource-based models attribute the processing asymmetry to differences in memory load posed by ORCs and SRCs. Semantic/pragmatic models emphasize that SRCs and ORCs differ with respect to their discourse-pragmatic and semantic properties, leading to a more straightforward interpretation of SRCs. Finally, frequency based models relate differences in processing difficulty to differences in distributional properties of the two sentence types.

Here, I will ignore semantic/pragmatic models, but rather use this phenomenon to flesh out the role of working memory vs. experience in accounts of human parsing. The opposition between memory-based and frequency or experience-based accounts does, of course, oversimplify the various differences between particular accounts, but can at least be used as a helpful demarcation line between accounts (cf. Staub, 2010).

2.1.2.1 Locality as a Determinant of Processing Complexity

The main idea underlying working-memory based models is that a central part of language processing relies on a resource with a limited capacity. While interpreting a sentence, elements have to be held in memory until they can be attached to their corresponding arguments to form a consistent structure. In section 2.1.1, I pointed out how ambiguity resolution has been explained by referring to working-memory as a limited capacity. Strategies that are efficient by keeping memory load low during parsing can in some cases result in being led down the garden-path and require reanalysis once the wrong trail is noticed. In the case of RCs that do not contain any obvious form of ambiguity, the problem is a different one. The general assumption is that dependent elements, e.g. heads and their arguments, have to be integrated to form a coherent representation of the sentential structure and its meaning. Since language processing is sequential by its very nature, the first of the dependent elements has to be held in memory or has to be (re-)activated/retrieved when the second element is encountered and is attempted to be integrated with the first one. In general, working-memory based models propose that processing differences stem from the different burden that the retrieval of the dependent element poses on working memory.

Gibson (1998, 2000) proposed an influential theory that argues along these lines. The *Dependency Locality Theory* (DLT) provides a word-by-word metric of processing complexity. Gibson (2000) assumes two major sources of complexity, conceptualized as two kinds of costs associated with processing: *storage costs* mirror the effort of keeping structure in memory, while *integration costs* reflect the process of integrating dependent elements. Crucially, integration costs incorporate the assumption that a processed element has to get reactivated whenever it has to be integrated with an element at a later stage in the sentence.

Integration costs apply whenever dependent elements (head and argument) have to be set in relation, and are – due to activation decay over time – a *function of distance* between dependent elements. *Storage costs*, on the other hand, occur when an element requires the prediction of a syntactic head that has to be held in memory - any open predicted head causes storage/memory cost.

Differences in both storage and integration costs have been shown to correspond with differences in word-by-word reading times in a wide range of complex sentences, (Grodner & Gibson, 2005; Gibson, 2000; Warren & Gibson, 2002), including the RC asymmetry. While in SRCs, the embedded verb “attacked” can be directly integrated with its argument, integration at the embedded verb in ORCs has to cross another discourse referent (“the reporter”), thus leading to higher integration costs in ORCs.

For the current purpose the crucial property of the account is that both components predict that working memory burden and the resulting differences in processing complexity are a function of distance or *locality* of the dependent elements, determined mainly by the number of new discourse referents intervening between them.

2.1.2.2 Similarity-Based Interference

Gordon et al. (2001), however, provided evidence that it is not merely distance and the number of discourse referents that intervene between two dependent elements that account for complexity. They showed that the *similarity* between the intervening elements and the element that has to be retrieved for integration plays a substantial role. The authors manipulated the match between the *noun phrase* (NP) that has to be integrated and the intervening NP. Processing difficulty was considerably higher when NPs are of the same type (e.g. full NP/full NP vs. proper Name/full NP) or if they share semantic properties such as animacy. Gordon et al. (2001) take their results as evidence for *similarity-based interference*. Difficulty in assessing dependent elements stems from interference between the elements to be retrieved and intervening materials. The more similar the intervening elements are to the element that is attempted to be retrieved, the more they compete, thus increasing the probability of retrieval failure and the time it takes to retrieve the correct element. The same idea has been picked up by Lewis and Vasishth (2005). They provide a detailed cognitive model based on ACT-R (Anderson & Lebiere, 1998; Anderson et al., 2004) that models similarity-based interference and complexity of long-distance dependencies by using the retrieval-dynamics provided by the chosen cognitive architecture. In this model, elements are attempted to be retrieved from memory by a number of retrieval cues (e.g. animacy, type, etc.). The more cues the intervening elements share with the correct element, the higher the probability of misretrieval and the longer it takes to retrieve an element. A number of empirical studies provide evidence that retrieval interference based on similarity between items provides a more adequate explanation for processing difficulty of long-distance dependencies than distance alone (Gordon, Hendrick, & Johnson, 2004; Gordon et al., 2006; Fedorenko, Piantadosi, & Gibson, 2012).

2.1.2.3 Anti-Locality Effects as Evidence for Predictive Processes

Both Gibson's (2000) DLT and interference-based accounts relate processing difficulty to processes of memory retrieval and, crucially, predict increasing processing difficulty if elements have to be integrated with more distant elements. Konieczny (2000), however, reported results that seem to show the exact opposite: In some circumstances, reading-data shows *anti-locality* effects. In an eye-tracking experiment, participants read sentences with center-embedded (12) or extraposed (13) relative clauses.

(12) Er hat das Buch, [das Lisa gestern gekauft hatte], hingelegt.
He has the book, that Lisa yesterday bought had, laid down.
"He has laid down the book that Lisa had bought yesterday."

(13) Er hat das Buch hingelegt, [das Lisa gestern gekauft hatte].
He has the book laid down, that Lisa yesterday bought had.
"He has laid down the book that Lisa had bought yesterday."

Locality-based accounts predict higher processing costs on the clause-final verb when a RC has to be crossed to integrate head and argument, as in example (12). However, the verb was read faster in this condition, indicating that processing is not necessarily made more difficult by intervening elements, but can sometimes be *facilitated*.

These results are in stark contrast to locality-based accounts, but do provide evidence for the *anticipation-hypothesis*. The intervening materials, although increasing the distance, provide a more restrictive context for the following elements, and thus allows the parser to anticipate upcoming elements, resulting in lower processing costs. In a subsequent paper, Konieczny and Doering (2003) provided further evidence for anti-locality effects in a reading study, and propose a simple-recurrent network model as an instance of sentence processing mechanisms that are based fundamentally on prediction.

Anti-locality effects have also been found in Hindi (Vasishth & Lewis, 2006). In contrast to the connectionist model of Konieczny and Doering (2003), however, the authors provide a cognitive model based on the cognitive architecture ACT-R (Anderson et al., 2004), that attempts to incorporate both effects of locality and anti-locality (loc: decay; anti-loc: reactivation), as well as effects of similarity on processing complexity.

Currently, it seems uncontroversial that both locality and expectability[4] play a role with regard to processing-complexity of sentences. However, given the data so far, it remains an open question whether the data can be best explained by memory-based accounts, as provided by Vasishth and Lewis (2006), or can be better incorporated into approaches that emphasize prediction as the fundamental building block of human sentence processing.

2.1.2.4 Individual Differences and Processing Complexity

Another area of research focusing on the role of working memory on sentence processing has been established in research on individual differences. King and Just (1991) as well as Just

[4]It should be noted that *locality* as a descriptive category does not provide an explanation per se. Thus, effects of distance could in principle result from processes of prediction.

and Carpenter (1992) provided evidence that differences in working memory capacity interact with the processing difficulty in RCs. Working memory capacity was assessed with the reading span test (Daneman & Carpenter, 1980), and participants in a reading study were assigned to span groups (high/mid/low) according to the test. Reading span did interact with reading time so that span had considerably higher influence when processing the more complex ORCs. Just and Carpenter (1992) argued that the interaction of reading span and sentence type indicates that ORCs pose higher demands on working memory, which induces processing difficulty if memory capacity is relatively low. Just and Carpenter (1992) provided a symbolic processing model based on a production rule system (CC-reader) with a limited capacity resource that is used to perform parsing actions. By manipulating the capacity of this resource, the authors were able to model the results from King and Just (1991) by showing an increase in processing time with lower capacity, and more so for ORCs.[5]

An experience-based alternative In 2002, MacDonald and Christiansen provided a radically different explanation for both the complexity difference and the interaction with reading span scores reported in Just and Carpenter (1992). In contrast to positing a working memory capacity, their explanation draws essentially on linguistic experience. On the one hand, they claim that differences in the reading span test do not necessarily reflect a fixed capacity that varies between individuals, but could rather reflect more efficient processing due to linguistic experience. Consequently, complexity differences are not ascribed to different memory load, but to different distributional properties of the structures incorporated in SRC vs. ORCs. While the basic argument structure "who (subject) attacked (verb) the senator (object)" in SRCs obeys the regular *subject-verb-object* structure in English main clauses, in ORCs this word-order is reversed to "who (object) the senator (subject) attacked (verb)". Thus, the SRC structure can easily be transferred from simple main clauses, but not the ORC structure. Crucially, to account for the interaction of reading-span (taken as reflecting linguistic experience) the authors refer to the well-known Frequency × Regularity interaction that has been reported in word reading (Seidenberg, 1985).

To support their claim, MacDonald and Christiansen (2002) provide a connectionist model. They trained simple-recurrent networks (see section 2.2.3.1 for a more detailed description of

[5]Both the results of King and Just (1991), as well as the operationalization of working memory capacity with the reading span task have been discussed controversially (Caplan & Waters, 1999). However, since individual differences are not the focus of this work, I will not review this debate here. For a discussion of the reading span test see, for example Conway et al. (2005).

the architecture) on sentences generated by a simplified grammar of English, where 5% of the sentences contained either a SRCs or an ORCs (or more than one, since the grammar allowed for recursive embedding). 10 corpora of 10000 sentences each where generated and 10 networks where trained for three epochs of 10000 sentences, where each network was trained on one of the corpora. As a measure of network performance that relates network output to behavioral data, the authors report the *Grammatical Prediction Error* (*GPE*). This measure compares the probability of each word at a particular position in the sentence, according to the underlying grammar, with the activation of the networks' output nodes. Since the networks are trained to predict the next word at each position in a sentence, the vector of output activation represents the expectations of the network at each position. Difference in linguistic experience was taken to be reflected by the different amount of training (1, 2 or 3 epochs). The pattern of *GPE* values strikingly resembles the reading time results from Just and Carpenter. The amount of training mostly affects the error rates in ORCs, but not in SRCs – revealing an interaction of RC-type and linguistic experience as the amount of exposure to sentences.

MacDonald and Christiansen (2002) thus provided an account that is crucially based on experience as the main factor that determines processing complexity. The probabilistic structure of the linguistic environment is addressed as the driving force of complexity, realized by the interaction of a rather general learning mechanism (the SRN) and the sequential structure of the linguistic environment. Their SRN model instantiates an approach of sentence-processing that is experience-based at its very core: The fundamental mechanism in SRNs to acquire language is prediction, and differences in processing complexity are explained as differences in predictability of the incoming element. As a consequence, the general approach provides a radically different view on the representation of linguistic knowledge and in particular on the explanatory value of working memory as a central concept. Further evidence for the role of experience in the processing of complex sentences has been provided in a training study reported in Wells, Christiansen, Race, Acheson, and MacDonald (2009).

False alarm predictions as a source of complexity Apart from the nice opposition to memory-based approaches, there is another reason for describing MacDonald and Christiansens' (2002) model here in detail. The task of an SRN is to predict the next word in a sequence. The accuracy of these predictions in relation to the grammatically possible elements at a particular position is measured as the output error. A higher error thus tells us that the network either did not activate the correct elements enough (*miss*), or activated elements that were not grammatical (*false-alarms*). Although the GPE takes both cases into account, the resulting measure does not reveal any information about what was actually predicted, in other words, what causes the error. The networks in MacDonald and Christiansen (2002), however, showed considerably high error rates even after three epochs of training, which raises two questions:

1. Given the high error rates, is it legitimate to say that the networks acquired the language at all?

2. What exactly causes the high error rates? What exactly do the networks predict?

Both points are particularly important since the simulations are proposed as cognitive models, explaining human behavior – thus, it does not suffice to show that error patterns resemble behavioral results. In contrast, it also implies the claim that the processes of the model resemble processes in human sentence processing.

A critical evaluation of these points can be found in Konieczny, Ruh, and Müller (2009), and I will point out one observation, since it has been one of the starting points with regard to research on local syntactic coherence within our lab and, thus, of the research program my thesis is a consequence of. In an attempt to replicate MacDonald and Christiansens' (2002) simulation results, Konieczny, Ruh, and Müller (2009) found out that the networks showed considerable *false-alarm activation* of globally incorrect elements. In particular in earlier epochs of training, the networks did activate incorrect elements considerably higher that grammatical ones, thus, with regard to the first question (1.), it calls into question whether the networks have been acquiring the crucial properties of the language at all. Consequently, it seems at least questionable whether the particular implementation provides convincing support for MacDonald and Christiansens' claims with regard to the relative clause asymmetry, since the networks do not provide a particularly good model of a competent language user. More interesting with regard to the topic of this work, however, is a different observation. Analysis of false-alarm activations revealed that the grammatically incorrect predictions are consistent with the *local* preceding context. In other words, the networks showed effects of *local syntactic coherence*. Thus, given that SRNs can be shown to be capable of acquiring a language to a reasonable degree, in contrast to the actual model provided by MacDonald and Christiansen (2002), a hypothesis that can be derived from the general framework is that predictions based on local context should provide a considerable source of processing complexity.

Effects of local coherence thus provide a distinctive prediction of experience-based models of sentence processing. More specifically, the effects are only expected if, in contrast to the *self-consistency* assumption, predictions are not only derived as continuations of a fully-connected structure that takes into account only grammatically possible continuations given the sentence input so far. In contrast, whatever sequential structures occur within the linguistic environment with sufficient frequency will, mediated by the process of sequential learning, contribute to the expectations about incoming materials.

2.1.2.5 Preliminary Discussion

Two main mechanisms have been reviewed with regard to processing complexity. On the one hand, effects of distance or locality between dependent elements have been taken as evidence for the central role of working memory associated retrieval processes. On the other hand, linguistic experience and, in particular, prediction have been put forward as either an alternative explanation for processing differences or, at least, as an additional factor that determines processing difficulty in unambiguous sentences, supported, for example, by effects of anti-locality which are hard to account for by working memory based approaches. Effects of predictability have been claimed to support connectionist approaches that provide radical experience-based accounts of sentence processing. In general, so far there's no conclusive evidence about whether effects of both locality and anti-locality can be accounted for within one mechanism, or if the effects mirror independent processes involved in sentence processing. As has been pointed out with regard to ambiguity resolution, the approaches provide rather different perspectives on cognitive mechanisms and processes in general. While working memory based approaches in general propose processes of retrieval and thus imply discrete representation of linguistic units, under a connectionist approach, representation and processing are not conceived of as a system of iterative rule application and memory retrieval. Rather, comprehension of a sentence is construed as a process of continuous dynamical change of the system state, guided by the probabilistic information entailed in the input, resulting in a position in the state space representing the interpretation of the sentence. A distinctive prediction of the latter approach, in particular with regard to the assumption of self-consistency, are effects of local syntactic coherence, as resulting from frequent occurrence of sequences in the linguistic environment, that can lead to globally incorrect predictions if such a sequence occurs in a different context.

In the following section, I will focus on usage, experience and prediction in sentence processing in more detail. A crucial distinction will be made with regard to approaches conceiving prediction explicitly as a *rational* strategy and, consequently, subscribe to the *self-consistency* assumption. As a counterpart I will describe connectionist approaches to sentence processing, in particular simple recurrent networks. Prediction, and therefore experience, is the basic factor that is claimed to form sentence processing in both approaches and, at least at some level of description, they bear crucial similarities. However, the self-consistency assumption implies a rather different concept of grammatical knowledge, and, in consequence, leads to different predictions with regard to sources of complexity.

2.2 Usage and Experience in Sentence Processing

2.2.1 Prediction and Sentence Comprehension

In the last section, I discussed the role of prediction – and thus experience and usage – with regard to both ambiguity resolution and processing complexity. However, prediction on the basis of knowledge derived from the distribution and associations of events is of course not only vivid in sentence processing, but is a fundamental property of (human) cognition in various domains. Whenever acting or reacting, humans anticipate and predict events, or at least have more or less specific expectations. In social interaction, the anticipation of the behavior of other agents is crucial, and social norms can be described as making behavior predictable. Control of attention is fundamentally influenced by the occurrence of unexpected events, as is evident in the influence of salience on perception. Finally, to violate the expectations of others plays an important role in specific human behavior, for example in humor. The possibility of violating expectations on purpose crucially requires, as a first step, anticipating expectations of others.

As Bar (2009) puts it, prediction can be termed as "using our past to prepare for the future." and puts forward the hypothesis that generating predictions could serve as a universal principle for understanding the operations in the human brain:

> Given exciting developments in theory, empirical findings and computational studies, it seems that the generation of predictions might be one strong candidate for such a universal principle. (Bar, 2009, p. 1181)

Predicting, anticipating and expecting something is intimately linked with experience. "Using our past" refers to this intimate connection with knowledge as the basis for prediction. While sometimes this knowledge may consist of the memory of a particular, unique situation, in most cases it will be more complicated. In the majority of circumstances, we don't know exactly what will happen, but rather maintain some weighted set of hypotheses, where the weights stem from repeated exposure to similar situations. In other words, predictions are based on accumulated knowledge which gives us the possibility to estimate the probability of possible events given a particular context.

As in cognition in general, experience, prediction and anticipation have been shown to be important factors in language processing.[6] A basic aspect of experience and "using the past" in language processing is the long-standing insight that more frequent words are read faster than less frequent ones (Rayner & Duffy, 1986). More recently, Arnon and Snider (2010) also provided evidence that not only word frequency influences reading-time, but also that frequent more-word sequences are read faster than infrequent ones. Effects of word frequency can be explained as resulting from easier accessibility and, depending on the underlying theory, faster retrieval from the mental lexicon. The same holds true for Arnon and Sniders' (2010) results, although they provide evidence that frequency information is not only stored for words, but also for larger units, which can be interpreted as questioning words as the basic representational format of linguistic knowledge. Effects of word or, more generally, unit frequency, regardless of the particular mechanism it is ascribed to, can be termed as a bottom-up effect, mainly driven by the unit itself.

Thus, although providing evidence that the processing system somehow keeps track of distributional properties of the language, it can be distinguished from prediction as crucially depending on some preceding context information. The fact that language processing is sensitive to predictability has been shown in several studies that showed that cloze-probabilities of a word[7] are a reliable predictor of reading times (Ehrlich & Rayner, 1981). A variety of studies using event related potentials (ERP) revealed that cloze-probability of words is also strongly related to the amplitude of the N400 response (for a review see Federmeier, 2007). In contrast to word frequency, these results provide evidence that context-dependent probabilities have to be tracked and mentally represented, and influence processing as a top-down process, taking into account (at least) the context of the currently processed sentence. Somewhat in between is the influence of transitional probabilities of a word given only the preceding word (bigram-probabilites), which have also been shown to reliably influence reading times (Boston, Hale, Vasishth, & Kliegl, 2011; Boston, Hale, Kliegl, Patil, & Vasishth, 2008).

[6] I will not discuss influences of frequency on language change, production and the shape of language itself, which is a major topic in related areas. For reviews, see, for example (Diessel, 2007; Ellis, 2002).

[7] The probability that a particular word is produced as a continuation of a given sentence fragment.

The various results from reading studies provide extensive evidence that the processing of a word is dependent on its *predictability*. These results can be interpreted such that the currently processed word is easier to integrate into the incrementally built sentence structure if it has a higher expectability on various contextual ranges. The results, however, do not necessarily imply that, during processing, comprehenders use their knowledge of the distributional properties of a language to actively *predict* what will come next.

It has been shown prominently in several visual-world experiments that comprehenders indeed do this. A series of studies has shown that comprehenders actively anticipate upcoming elements before they have been mentioned in the sentence, and can make use of a variety of informational sources to do this. Altmann and Kamide (1999) showed that participants can use a verb's selectional restrictions to fixate visual depictions of a plausible argument even before the corresponding word has been heard. Similar experiments showed that comprehenders can make use of a variety of cues for anticipation, including case-marking (Kamide, Scheepers, & Altmann, 2003), discourse information (Kaiser & Trueswell, 2004), subcategorization information of verbs and relative frequency of verb-usage (Arai & Keller, 2012).

A variety of models of sentence processing has incorporated the influence of probabilistic information as a determinant of processing complexity and of initial preferences in ambiguous sentences. In probabilistic grammar-based models of sentence processing (Jurafsky, 1996), phrase structure rules are associated with probabilities. As mentioned before, interactive constraint-based approaches rely on the assumption that various probabilistic constraints determine the most likely interpretation at a current processing step, and dynamical models of sentence processing are fundamentally driven by their internal representations of the probabilistic properties of their (linguistic and extralinguistic) environment.

Several recent proposals implement models focusing explicitly on prediction as the fundamental mechanism (Demberg-Winterfors, 2010). I will not discuss the several models proposed here. Rather, I will describe two formalisms that have been proposed as rather abstract, broad coverage accounts of processing complexity as syntactic predictability. In contrast to cloze-probabilities, which are based on actual prediction, and thus, production, both accounts essentially draw on corpus data as representing, at least partially, the experience of a comprehender.[8]

[8] For recent approaches comparing cloze-probabilities and corpus-derived predictability, see Smith and Levy (2011) and Hahn (2012).

2.2.2 Prediction as a Rational Strategy

2.2.2.1 Surprisal

Surprisal as a predictor of required processing effort has been first proposed by Hale (2001), and taken up by Levy (2008a) as a incremental measure of processing difficulty representing the amount of resource allocation a word in context requires. The basic idea underlying the notion and implementation of syntactic surprisal, as Levy (2008a) construes it, is that the system allocates its resources to the set of structures compatible with the input processed so far, according to the probability each structure is associated with the input string. At each incoming word w_n, the probability distribution changes, and resources have to be reallocated accordingly. The amount of change that is necessary to arrive from the distribution having processed w_0 to w_{n-1} to the distribution over structures compatible with w_0 to w_n is assumed to reflect the cognitive-effort induced by w_n, and can be calculated as the negative log of the transitional probability of word w given the preceding sentence context:

$$surprisal_{w_i} = -log(P_{w_i}|P_{w_1...w_{i-1}}) \tag{2.1}$$

Levy (2008a) showed that the surprisal theory can account for a variety of established psycholinguistic phenomena, for example, gardenpathing. Crucially, he can also account for some phenomena where memory-based theories face difficulties, including anti-locality effects and facilitating effects of ambiguity that have been observed in particular contexts. Phenomena that pose difficulties for the approach are the subject-object asymmetry in relative clauses (section 2.1.2) where surprisal differs from observed data with regard to the time course of difficulty within the RC.

Although surprisal does not readily provide a particular implemented processing model, it nevertheless subscribes to a highly interactive and parallel system and thus, bears strong resemblance with constraint-based interactive approaches. However, it ascribes processing difficulty and elevated reading time not to competition between different alternatives, but to the requirement of redistributing attentional resources. In addition, surprisal implies a self-consistent, fully-incremental mechanism, since it is based on transitional probabilities regarding the current sentence. The context that is taken into consideration starts at the beginning of the actual sentence.[9] Taking into account the whole sentence so far can be considered as a

[9] Levy (2008a) explicitly states that extra-sentential context has also to be considered. However, in the context of this paper this can be ignored.

rational strategy for a parser. With every element, the search space is narrowed down, based on probabilistic knowledge about the language, so that at every point in time all and only potentially relevant information is used to update the expectational space.

In addition to the account for experimentally observed differences in reading time, evidence for surprisal as a significant predictor of different reading time measures has been provided by Boston et al. (2008) using the Potsdam reading corpus. Others, however, have provided evidence that surprisal and processing costs derived from memory-based approaches (as, for example, DLT) influence reading times independently, suggesting that predictability and working-memory load may stem from separate processes and must both be considered in models of language processing (Demberg & Keller, 2008; Boston et al., 2011).

2.2.2.2 Entropy Reduction

While surprisal represents the predictability of a particular word given the preceding context, the entropy reduction hypothesis (Hale, 2003, 2006) takes a slightly different perspective. Similar to surprisal, entropy reduction provides a word-by-word complexity metric that can be conceived of as the influence of a particular word on a comprehender's beliefs about the structure of the sentence. However, in contrast to surprisal, it considers not only the predictability, or rather, unexpectedness of a word given its preceding context, but complete analyses that are compatible with the already processed input words, amounting to assigning probabilities to all structures that a probabilistic grammar of the language can generate. The underlying idea is to conceive a word as reducing (or not) the uncertainty about the rest of the sentence – in other words, it reduces the entropy, as an information-theoretic measure of disorder, of the probability distribution over grammatical parses. Reducing entropy, or producing order, is assumed to require cognitive effort.

In two recent papers, entropy reduction has been shown to be a significant predictor of reading-time (Roark, Bachrach, Cardenas, & Pallier, 2009; Frank, 2010). In particular, entropy reduction can account for particular phenomena in relative clauses where surprisal is not compatible with empirical results (Chen, Jäger, & Hale, 2012).

As Hale (2011, p. 400) points out, both surprisal and entropy-reduction merely remain on the *computational* levels insofar as they describe a function relating properties of the input with output properties (for example, processing time). In other words, they provide useful characterizations of how sentence complexity can be measured, but remain relatively silent about the particular processes that realize this particular input/output relation.

2.2.3 Connectionism as a Cognitive Paradigm of Experience-Based Sentence Processing

The connectionist approach to cognition has been put forward as a radical alternative to the classical, algorithmic approach to cognition. The basic claim of the latter view is that a cognitive system best understood as a system that manipulates discrete symbolic representation according to formal rules; that the mind can be described as a digital computer, and the brain as the device that implements the mind's programs. In contrast, connectionism typically claims that cognitive processes can best be conceived of as emerging from parallel processes of highly interconnected, relatively homogeneous units. In contrast to symbolic representations, as, for example, lexical entries in a "mental lexicon", units in a connectionist system, an artificial neural network, normally do not directly represent symbols that "stand in" for entities in the world, but rather that the joint activity of multiple units provides the system with the means to react appropriately in its environment. It goes without saying that this blunt characterization can neither do justice to any of the two theoretical frameworks, nor includes the various variants of hybrid approaches that have been proposed, trying to combine the characteristics of both frameworks. Since detailed characterizations of both accounts, their differences and respective shortcomings, have been done elsewhere in detail, I will not attempt to go in any details here. Rather, I will concentrate on Simple Recurrent Networks (SRNs) as one particular connectionist architecture that has been highly influential with regard to language processing, and consider models based on this architecture as a particular instance of a radically experience-based account of both language acquisition and processing, as has already been pointed out as an example in section 2.1.2.4.

2.2.3.1 Simple Recurrent Networks (SRN)

A variety of different connectionist architectures has been applied to model different aspects of language processing, acquisition and change (for reviews see Christiansen & Chater, 1999a). Arguably the most influential architecture with regard to both language acquisition and sentence processing are simple recurrent networks (SRNs). Elman (1990) proposed this architecture as a relatively simple way to provide a connectionist architecture that is able to handle an essential property of language input – its sequential nature. The crucial insight of Elman in this paper was that language processing (and acquisition) has to be conceived as "finding structure in time". Elman introduced SRNs as a rather simple extension of a classical multi-

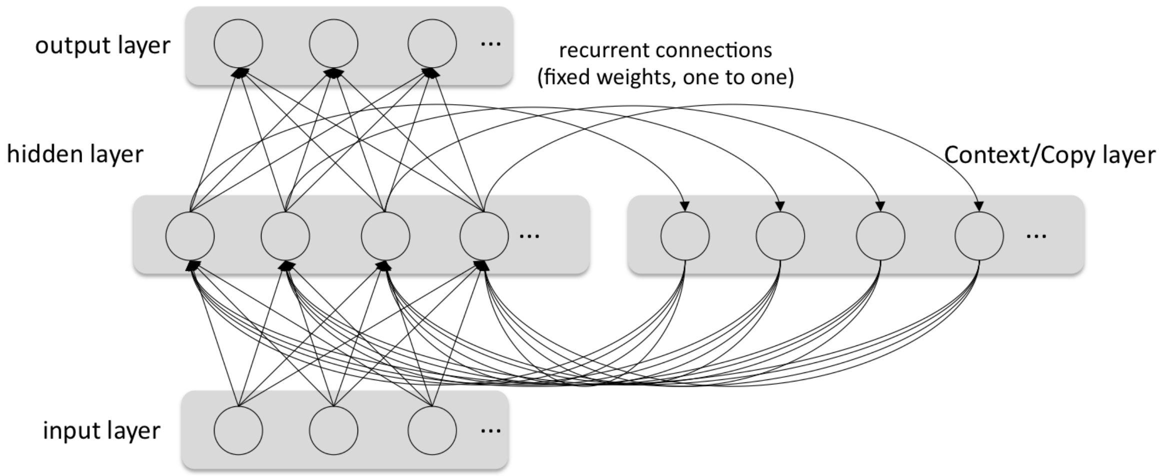


Figure 2.1: A Simple recurrent network

layer feedforward network, with the goal of taking into account the aspect of time and its role in cognitive processing. Rather than including time explicitly in a model, by, for example, adding an additional dimension that represents the current point in time, Elman attempted to model time implicitly, that is, by its effect on processing – which essentially means that the model has to be provided with some kind of "memory" that allows the model to keep track of preceding input. The very simple but successful and influential way Elman achieved this was by connecting an additional hidden layer to a feedforward network by recurrent connections. Figure 2.1 shows a small SRN. The left part (input, hidden and output layer) is identical to a classical feedforward network, where each node is connected to all the nodes of the next layer. However, the hidden layer is now connected to the additional layer, which we will refer to as the *context layer*. Three properties of the context layer are essential. Each node of the hidden layer is connected to not all but only one node of the context layer. In addition, the connection weights between hidden and context layer are all set to 1, and are not changed during learning. Finally, the activation function of the nodes in the context layer is linear. These settings ensure that the vector of activation of the context layer in time step t is an exact copy of the activation vector of the hidden layer in the preceding time step t-1. The context layer is connected to the hidden layer in the usual way (each context node is connected to all hidden nodes), thus providing *recurrent* connections.

At each time step, the input of the hidden layer is not only the current input, but, in addition, its own (weighted) activation vector from time step *t-1*. Thus, the network has access to its own internal representation of preceding input. Crucially, however, the input from the context layer is not just the internal representation of the input on time *t-1*, because the internal representation has already been a combination of the external *and* the internal input. So, the representation of the hidden nodes is a function of the external input and the complete past of the network, or in other words, of the complete sequence of external inputs the network received up to time point *t*. The central accomplishment of the architecture is that a particular input can be processed with regard to its preceding context. The task such a network typically has to achieve is to simply predict the next element at each step based on the sequential input it receives. The input of the network, thus, is a sequence of input signals, each representing, for example, a word of a sentence. The task of the network is just to predict the next word of the sequence. The produced output at each step is compared to the actual next element – the target or teaching signal, and the resulting error is backpropagated through the networks in order to change the connection weights.

Thus, the sequence of teaching signals is just the sequence of input signals shifted by one position. Given reasonably complex language underlying the input sequence, however, the network will not be able to achieve perfect performance, given the, in principle, infinite number of sentences. Using Elman's examples, the network will not be able to predict if the sequence "boy chases ..." will be continued by "cat" or "dog". However, what it can learn is to activate only the set of elements that are appropriate as continuations, which are, in our example, nouns. In other words, it can learn the categories that are licensed to occur after the sequence "boy chases", determined by the grammar underlying the training materials. In several papers, Elman showed that this simple architecture does remarkably well on extracting regularities by just predicting the next element. The notable property of SRNs is that they can extract categories and dependencies between distant elements just by being provided with proper sequences. In other words, they can learn the grammar that is underlying the training materials or, at least, crucial aspects of it, without being provided with any explicit representation of grammatical rules or categories. All the information the network can use to extract the relevant dependencies lies in the sequential order of the input signal and, crucially, the distributional properties that the input entails.

SRNs and the prediction task have been used in a broad range of studies, ranging from small scale studies exploring the capabilities of the architecture (Christiansen & Chater, 1999b; Tabor, Cho, & Szkudlarek, 2012) to large scale studies trying to cover a broad range of psycholinguistic behaviors (Rohde, 2002). A variety of studies successfully used (sometimes extended and modified variations of) SRNs to derive predictors of reading times (Tabor et al., 1997; Frank, 2009; MacDonald & Christiansen, 2002) as well as eyemovements in visual-world data (Mayberry III., Crocker, & Knoeferle, 2005; Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Kukona & Tabor, 2011; Weldle, 2011).

2.2.3.2 Analyzing SRNs

One issue that has been raised with regard to the explanatory value of connectionist models, including SRNs, is that it is not easy to assess how a network does what it does, which is, however, what we want to get back from a cognitive model of a particular cognitive ability or task. An important step to understand how an SRN can achieve its task is to analyze the internal representations that are developed during the training process. Several methods have been proposed to "look inside" a network, for example by cluster analysis of the activation patterns of the hidden-layer (Elman, 1990), or principal component analysis to visualize the network's trajectory over time through its state space (Elman, 1993).

A particularly fruitful approach to both analyze SRNs and provide the means to understand what it does is to view an SRN as a dynamical system. As such, the state of a network at each point in time is described by the activation values of its nodes, constituting a particular point in a multidimensional *state space*, which is defined by the set of possible values the nodes can take. Any input (internal or external) the network receives changes the system state. The state change over time, that is, the sequence of points the network goes through over time, provides the *trajectory* of the system. Crucially, the state space of a system is typically not uniform, but can be conceived of as a landscape with "hills" (*repellors*) and "valleys" *attractors*. Starting at a particular system state, the system's trajectory will tend to move downhill to finally settle at or near an attractor. Methods from dynamical system theory have been used, for example, to analyze how SRNs process recursive sentence structures (Rodrigruez & Elman, 1999; Tabor, Cho, & Szkudlarek, 2013).

In addition to the question of what happens "inside" the network, evaluation of performance itself is not always easy, as has been exemplified in section 2.1.2.4. I will come back to this issue with regard to the SRN simulation provided in this work in chapter 4.2.

2.2.3.3 Problematic Aspects of SRN Models

Free parameters A first point is the relatively high number of parameters that have to be chosen in order to provide a model that is capable of doing the task at all. Decisions on size and number of hidden-layers, learning parameters and initialization of connection weights have been shown to be highly influential with regard to performance of the network, and can be determine whether the model learns the task at all (Rohde & Plaut, 1999). The same holds true for the particular learning algorithm that is used to adapt the connection weights.

The amount of training, the structure of the training materials, and the particular procedure of training are further components that influence behavior. Elman (1993), for example, found out that particular generalizations were only learned by an SRN when it was first provided with a subset of less complex structures, and only with the full set in later stages of training (see, however Rohde & Plaut, 1999, for a critical evaluation of the results).

Representational format of input and output A crucial decision also regards the representational format of input and output. In most SRN applications regarding sentence processing, different words of the language to be learned are represented as orthogonal bit vectors such that for each word one input node is activated, while all others are set to zero. On the one hand, this ensures that the SRN can use only sequential information, and makes interpretation of the output vector considerably easier. On the other hand, this kind of localist input and output representation, although deliberately, ignores potentially relevant similarities between individual words in a language. In addition, since each token of the language requires one input and output node, localist representation significantly narrows down the vocabulary size that can reasonably be used in the model. This latter issue points to a question that is often raised with regard to SRN models. Since most models rely on a very limited vocabulary and on rather simplified grammars, mainly including the specific features that are of interest, the question arises as to how good such models can serve as general models of sentence processing. The whole complexity and richness of linguistics forms is not captured, and thus, it is hard to know if SRNs provide a good model for "real language" processing. However, several proposals have been attempted to address this issue by using more elaborated, distributed representational formats and to train SRNs on, at least parts of, real language corpora (Frank, 2009, 2013).

Link to behavioral data Finally, the output of a classical SRN is nothing more than a vector of activation values for each word in the language it has been trained on, which can be conceived as expecting each word with a particular probability. Although prediction is arguably a rather fundamental ingredient of human language processing, the output of the SRN can not easily be matched to any behavioral response of a human comprehender. In other words, the link between model outcome and human data is rather indirect. Several methods have been proposed both to assess performance of SRNs, and to provide a linking function between activation values and, for example, reading times (MacDonald & Christiansen, 2002). A variety of approaches has used extended variants of the classical architecture to provide a closer link between model outcome and experimental measures (see Rohde, 2002; Mayberry III. et al., 2005; Kukona & Tabor, 2011; Weldle, 2011; Tabor et al., 1997, for examples).

The property of a standard SRN to provide an output activation vector of all elements in the grammar which can be treated as a probability distribution over possible continuations, provides relatively easy ways of comparing outcomes of a SRN with the prediction-based measures as surprisal and, at least partially, entropy reduction (Frank, 2009, 2013). The crucial difference, as has been pointed out above, is that the prediction of SRNs, although trained on materials generated by a "classical" rule-based grammar, do not strictly reproduce the distributional properties, as it is the case if surprisal or entropy reduction are calculated from a probabilistic context free grammar, or, as in Chen et al. (2012), from a minimalist grammar. Thus, while the underlying grammars may be identical, the output will probably not. In some sense, this of course indicates that the network did not learn the grammar correctly. However, as Frank (2009, p. 82) points out,

> An accurate language model is not necessarily an accurate psycholinguistic model. [...] this is because the extent to which words are surprising to a reader may deviate from the surprisal values as estimated by the language model.

Thus, in terms of rationality, the suboptimal use of distributional properties reflected in the SRN output, which is, for example, evident in false-alarm predictions due to local syntactic coherence (see paragraph 2.1.2.4), may reflect similar "irrational" aspects in human behavior.

To conclude, SRNs proved, on the one hand, a powerful and relatively simple model of some aspects of human language acquisition as implicit grammar learning based on positive input, and as models of language processing that can account for a variety of linguistic phenomena with relatively few presuppositions. On the other hand, a number of questions remain open, some of which will have to considered again in subsequent sections of this work.

2.3 Local Syntactic Coherence: Evidence and Theories

In the previous sections we saw that there's ample evidence to show that predicting upcoming elements and maintaining expectations about subsequent elements is an important mechanism in sentence processing. This insight is incorporated in virtually all current models of human language processing, and is, as we have seen, at the very core of connectionist approaches that are based on simple recurrent networks. I also pointed out that the question about the particular context and and the different sources of information that are used to build up expectations and derive predictions are a central ingredient of theories on human language processing. A large corpus of research shows that comprehenders use a wide range of sources to quickly derive the structure and meaning of a sentence or utterance, and can make use of subtle hints in the input to predict probable continuations.

Most theories assume that comprehenders use the rich sources of information to narrow down the set of possible structures and interpretations as the sentence unfolds, and build up partial, but consistent representations of the information provided. Research in ambiguity resolution provides evidence that the available information can sometimes lead a comprehender down the wrong track, requiring reanalysis or cognitive effort to adapt a currently supported "guess" to subsequent input. Although theories differ in their assumptions about which information sources come into play at what stage in the process, in general they share the assumption of incremental, immediate and self-consistent parsing. These assumptions are made explicit in principle-based theories of ambiguity resolution such as the Garden-Path model, and underlie the attempt to construe parsing as a "rational" process. Self-consistent parsing thus implies that the parser in general does not *ignore* available information that could be used to constrain the set of candidates of possible syntactic structures of a sequence of words, since such "ignorance" would be hard to account for given the requirement of efficiency in processing.[10]

Errors or difficulties due to temporal or global ambiguity, are direct consequences of the incrementality and immediacy of parsing, and are compatible with the self-consistency assumption. If multiple grammatically correct alternatives are possible at a particular stage, no available information can prevent the parser from running into difficulties or errors. Although it is still controversial which information is available at a particular stage in the parsing pro-

[10]Exceptions from self-consistency could be seen in cases of multiply center-embedded sentences, where erroneous interpretations or judgements have been attributed to working-memory limitations, assuming that not all available information can be taken into account because it exceeds working-memory capacity (Gibson & Thomas, 1999).

cess, it is rather uncontroversial that the available information is not ignored. Effects of local syntactic coherence have been identified as evidence that bring the self-consistency assumption in question. Although, in some sense, compatible with immediacy and incrementality, in some cases the human language processing system seems to build up structures that are not consistent with the preceding sentence context. Thus, predictions do not necessarily seem to be completely constrained by the sentence context, and expectability of elements may be influenced by local context, regardless of information that is already available. A local substring may be associated, based on its surface form, with a syntactic structure that can not be integrated with the preceding context. In other words, in some cases the human parser builds up ungrammatical structures.

I will now first introduce and define the phenomenon of local syntactic coherence in more detail and provide a number of examples to give an impression of different cases of its various occurrences. Following that, I will report the data that has been taken as evidence for the processing of local syntactic coherence so far, and the various attempts to explain these results.

2.3.1 Defining Local Syntactic Coherence

Tabor et al. (2004), who were the first to report measurable influences of *local syntactic coherence* on sentence processing, define local syntactic coherence as follows (p. 355):[11]

> By 'local syntactic coherence in the input' we mean sequences of two or more words in the text stream which form a phrase or part of a phrase that cannot be grammatically unified with the parse of the preceding words in the sentence.

Bicknell et al. (2009, p. 6) give a similar, but slightly broader definition, including also one-word local coherence:

> We take the definition of a local coherence to be a string of words w that out of context would suggest one very likely parse, and that parse is impossible (or at least highly unlikely) in context.

Even broader, local coherence can be defined as induced by any sequence of linguistic elements of arbitrary length that can occur in at least one function or interpretation that is not consistent with the preceding context it is embedded in. Thus, on the level of strings of words, the most basic local syntactic coherence is constituted by any ambiguity of a word, such as

[11] Note that Bicknell et al. (2009) do, in contrast to Tabor et al. (2004) not refer to local *syntactic* coherence.

noun/verb ambiguity, case ambiguity or, as we shall see in the following, morphological ambiguity between adverbial participle and finite verb.[12] In principle, local coherence can be described at any level of linguistic description, ranging from sequences of phonemes or letters up to sequences of sentences or utterances. Allopenna, Magnuson, and Tanenhaus (1998), for example, showed in a visual-world experiment that participants listening to words like "beaker" fixated more on rhyme competitors ("speaker") than on distractors with a different word onset, which can be seen as evidence for effects of local coherence on the sublexical level.

Regarding one-word local coherence, although not referred to under this name, several studies showed that multiple meanings of ambiguous words are activated when lexically ambiguous word are processed, even if the preceding syntactic or semantic context strongly favors one particular meaning (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). However, effects of one-word local coherence could be attributed to bottom-up processes of an independent lexical module, and do not readily call into question self-consistent parsing on the syntactic level: Although an incompatible meaning may be retrieved, it is not necessarily integrated with any of the elements that have been processed beforehand. Rather, a self-consistent parsing mechanism should select the consistent meaning based on the syntactic and semantic requirements of the already processed input. A wide range of theories has been proposed about the processes and the time-course of lexical access, which can not be described within the scope of this work. Influence of what can be counted as of one-word local coherence due to lexical category ambiguity would be expected under a rather broad range of theories relying on very different theoretical frameworks. More-word local coherence, however, provides a more distinctive testbed, as it seems to pose severe problems for self-consistent parsing approaches. Thus, it also seems to call into question the assumption that human sentence processing can, in general, be conceived of as a rational, optimally adapted, device for the task at hand. Therefore, I will focus on local syntactic coherence on the phrasal level, subscribing to the more restrictive definition of local *syntactic* coherence as a sequence of (more than one) words that, viewed in isolation, have an interpretation that is not consistent with the sentential context they're embedded in.

Two-word local syntactically coherent structures abound in language, since basically all noun phrases, in particular in languages with poor case-marking, can occur in several grammatical functions and thematic roles that may differ from the one they take in the context of a

[12] Note however, that local syntactic coherence does not necessarily involve ambiguity at the lexical level (Kukona & Tabor, 2011; Weldle, 2011)

particular sentence. More interesting, however, are cases where the locally most likely syntactic structure of a sequence of words can be said to express a *proposition* that is not expressed by the global sentence. In the following, I will give some examples where this is the case. In all these examples, the local syntactic coherence is induced by a string of words that can form a complete main clause (marked by curly brackets).

(14) The coach chided {the player tossed a frisbee} by the opposing team.
(Tabor et al., 2004)

Considering this sentence, it is the player that receives a frisbee, while the local sequence represents the player as tossing the frisbee himself. Local syntactic coherence in this case is induced by the morphologically ambiguous *tossed*. In German, similar structures are possible due to a range of morphological ambiguities, for example using verbforms that can occur both as a finite verb and adverbial participle, such as «begeistert» [enthuses/enthusiastically] in example (15), or as past-tense verb and adjective, as «belegte» [filled] in example (16).

(15) Peter zeigt {Maria begeistert$_{ambig}$ den Clown}.
Peter shows {Maria enthusiastically/enthuses the Clown}.
"Peter enthusiastically shows the Clown to Maria."

(16) Als {der Gastgeber belegte$_{ambig}$ Brötchen mit Käse} servierte, freuten sich
When {the host filled rolls with cheese} served, pleased
die Gäste.
were the guests.
"When the host served filled rolls with cheese the guests were pleased."

However, local syntactic coherence does not necessarily depend on the presence of a lexically or morphologically ambiguous verbal element (Weldle, 2011):

(17) Als der Vater, den {der Sohn kämmt, sich im Wohnzimmer} anzieht, ist die
While the father, that {the son combs, himself in the living room} dresses, is the
halbe Stadt bereits auf dem Weg zur Aufführung.
half town already on their way to the show.
"While the father who the son combs dresses himself in the living room, half the town is already on their way to the show."

In this case, the local string «der Sohn kämmt sich im Wohnzimmer» [the son combs himself in the living room] implies reflexive use of the verb, while globally it is used as a transitive verb with the son being the agent and the father the patient of the combing. As in the stimuli of Tabor et al. (2004), in this case local syntactic coherence is only induced by crossing a sentence boundary, in contrast to examples (15) and (16), where local syntactic coherence is completely embedded within the subordinate clauses.

The sentence

(18) Als der Postbote, den {die Floristin beschenkte, den Müllmann
When the postman, that {the florist presented [with something], the garbageman
auf der Straße} traf, wurde es gerade dunkel.
in the street} met, got it just dark.
"When the postman to whom the florist gave a present met the garbageman in the street it was getting dark."

from Konieczny, Müller, Baumann, Hachmann, and Wolfer (2009) is another example of a boundary-crossing local syntactic coherence, without including lexically ambiguous forms. Here, the local string implies a different recipient of the RC-verb «beschenkte» [gave a present to someone], namely «den Müllmann» [the garbageman], while globally the «der Postbote» [the postman] has to be considered as the recipient.
Local syntactic coherent sequences can even consist of complex clauses and can carry on until the end of the entire sentence:

(19) Psycholinguists reviewing {recent papers on human language processing report growing evidence for effects of local syntactic coherent sequences on sentence comprehension.}

While all the above examples are manually constructed, whole-sentence local syntactic coherence does also occur in natural texts. A particularly nice example can be found in a German newspaper article about a former German minister accused of having been involved in house occupation and violent activities in the 1960s.

> Was denn mit 'Putzgruppe' gemeint sei, möchte die Kammer wissen. Darunter sei der Versuch zu verstehen, gegen {geschlossen anrückende Polizei besetzte Häuser} zu verteidigen, antwortet der Zeuge. 'Aber wir sind jämmerlich gescheitert'. (Mannheimer Morgen, 22.03.2006; Als Zeuge gibt sich Joschka Fischer äußerst vorsichtig).[13]

[13] Rough translation: The court asks what should be meant by 'Putztruppe' [group of cleaning personell]. This meant the attempt to defend occupied houses against massively approaching police, the witness answered. 'But we failed miserably."'

While the paragraph is about the attempt to defend occupied houses against the police, the local string «geschlossen anrückende Polizei besetzte Häuser» expresses that the police itself occupied houses [massively approaching police occupied houses].

In all the above examples, a sentence contains a string that forms a complete simple or even complex sentence, but the preceding context rules out a main clause interpretation of the sequence, so that there is no structural ambiguity in the classical sense to be found anywhere in the sentence. Hence, a parsing mechanism strictly operating in a self-consistent way as described above should not predict any effect if a sentence contains a locally coherent sequence. Taking, as an example, surprisal-based accounts, the set of possible syntactic structures at the point where the ambiguous element is encountered, would not include any structure containing the locally coherent, but globally incorrect interpretation of the ambiguous element. Thus, given the context, no difference of processing costs or extra cognitive load would be predicted comparing a sentence of this kind and a similar sentence without a local syntactic coherence.

In the next section, I will report and discuss the empirical results that have been taken as evidence for effects of local syntactic coherence, and the theories and claims that have been proposed to account for these results.

2.3.2 First Evidence and Models

2.3.2.1 Self-Organized Parsing

Tabor et al. (2004) were the first to have claimed that local syntactic coherence can influence sentence processing. In a series of self-paced reading experiments participants read sentences either containing a local syntactically coherent sequence (20a) or similar control sentences without local syntactic coherence (examples (20b) to (20d)).

(20) a. The coach smiled at {the player *tossed* a frisbee} by the opposing team.
b. The coach smiled at the player who was *tossed* a frisbee by the opposing team.
c. The coach smiled at the player *thrown* a frisbee by the opposing team.
d. The coach smiled at the player who was *thrown* a frisbee by the opposing team.

Local syntactic coherence in example (20a) is induced by the ambiguous element *tossed* (past participle and past tense form), leading to the word sequence {the player *tossed* a frisbee}, which is identical with a simple main clause with *tossed* as its past tense verb. Given the preceding context, however, *tossed* has to be interpreted as a past participle, since it constitutes the verb of a (reduced) relative clause in passive voice. In the three control conditions, local syntactic coherence is inhibited either by using verbs with an unambiguous passive form in (20c) and (20d), or by using the unreduced relative clause in (20b) and (20d).

In addition to the expected main effect of relative clause type (reduced RCs are read longer than unreduced RCs), Tabor et al. (2004) found higher reading-times on the locally coherent region in example (20a). The interaction between RC type and ambiguity is taken as evidence that the local substring is processed with *tossed* as a past tense verb, and that the resulting subparse competes with the globally correct interpretation.

As Tabor et al. (2004) note, the difference between ambiguous and unambiguous conditions could also be explained as a result of the – context-independent – lexical bias of the ambiguous/unambiguous verb. The conflict between syntactic requirements and lexical properties of the ambiguous element would then result in a slowing down of processing. Given the additional assumption that the influence of a more constraining context, as in the unreduced relative clauses, minimizes the influence of lexical properties, such an approach could also account for the observed interaction. This kind of explanation would thus ascribe the effect not to competition between local and global subparses, but rather to the interaction of processes of a lexical module and the syntactic processor.

The second experiment in Tabor et al. (2004) attempts to rule out this account and, in addition, tests if semantic properties of the local substring influence processing. This second aspect merely aims at distinguishing different accounts of self-organized parsing that differ with respect to the format that is assumed for the intermediate representations of sentence structure. In their second experiment, Tabor et al. (2004) used sentences that did not differ with respect to the verb, but included animate and inanimate nouns to manipulate semantic coherence of the local subsequence.

(21) The bandit worried about . . .

- a. {the prisoner transported the whole way}
- b. the prisoner who was transported the whole way
- c. the gold transported the whole way
- d. the gold that was transported the whole way

. . . by the capricious guard.

Again, longer reading times where observed at the locally coherent region in example (21a), providing further evidence that local syntactic coherence can induce processing difficulty. In addition, the influence of animacy indicates a) that the effect can not readily be explained solely by context-independent lexical processes, but depend on a broader local context, and b) that the effects are not restricted to a purely structural level, but incorporate semantic properties of the input string as well.

Tabor et al. (2004) take their results as strong evidence for the self-organizing structure of the human language processing system. Their own model is a self-organizing interactive approach and models the effects as competition between a local subparse and the globally correct one. The SOPARSE (Self-Organized Parser) model[14] is strongly inspired by the Unification Space approach proposed in Vosse and Kempen (2000). The atoms of the model are lexically anchored tree fragments ("strands" or "trelets"), consisting of a head node, a lexical anchor and foot nodes. A perceived word activates its tree fragments which forms links with other activated tree fragments. The links are weighted and their weight is dynamically changing, competing with other links to the same head- or foot nodes. The next word is perceived when the activation dynamics for a particular word have stabilized. Processing time is modeled as the cycles it takes the system to settle in a stable state. In the case of a local syntactic coherence, the local substring builds up a partial tree that competes with the globally correct structure, so that the system needs more time/cycles to settle. The explanation provided by this model is basically a process of interference of several competing structures/interpretations of different scope that are pursued in parallel, in our case the globally correct and the only locally coherent substructure.

Some criticisms can be raised both regarding the empirical results and the proposed model. With regard to the experiment, in particular experiment 1, one problem can be seen in the use of rather rare and complex sentence structures. As is evident from an additional grammaticality judgment study (Tabor et al., 2004, experiment 3, pp. 364), the reduced/ambiguous sentences were rather often judged as ungrammatical. Although this can be taken as further evidence for the disturbance induced by the local syntactic coherence, also the very low rate of *yes* judgments in the reduced/unambiguous cases at least provide evidence that processing of the stimuli was rather hard if not impossible in some cases. Thus, it can be called into question whether the observed effects can be seen as reflecting processes present in normal reading, or rather as consequences of processing breakdown triggering particular strategies to extract at least some meaning (this point is discussed in more detail in Tabor et al., 2004, p. 365). Regarding the SOPARSE model, one criticism that has been raised in Bicknell et al.

[14]The model is described in detail in appendix A in Tabor and Hutchins (2004, pp. 445-448).

(2009) is that the model's results depend on a rather high number of free parameters which are hard to legitimate independently.

Overall, although both empirical data and the proposed model leave a number of questions open, the paper made an important step in identifying a phenomenon that has not been considered before, and that provides an important test case to prove some of the basic assumptions that have been put forward with regard to the general principles the language processing system is operating on. Consequently, subsequent research has attempted to either provide further evidence for the phenomenon, and/or to explain the reported results by different means.

2.3.2.2 Locally Coherent "False Alarm" Activation

Evidence for effects of local syntactic coherence was also provided in Konieczny (2005) in an online anomaly detection experiment, which was motivated by the observation that in simple recurrent networks the activation of locally induced, but grammatically illegitimate elements provide a considerable source of error, already discussed in section 2.1.2.4 (see also Konieczny & Müller, 2010; Konieczny, Ruh, & Müller, 2009). To test if such *false alarm* activations of ungrammatical elements also occur in human language processing, participants were presented with subject- and object-extracted relative clauses which where rendered ungrammatical by adding an additional noun phrase after the RC object (in SRCs) or RC subject (in ORCs) (example (22)).

(22) Der Abgeordnete, ...
The congressman$_{NOM}$, ...

a. der den Journalisten beschimpft *[den Politiker]
who$_{NOM}$ the$_{ACC}$ journalist$_{ACC}$ attacks *[the$_{ACC}$ politician]

b. der {den Journalisten beschimpft *[der Politiker]}
who$_{NOM}$ the$_{ACC}$ journalist$_{ACC}$ attacks *[the$_{NOM}$ politician]

c. den {der Journalist beschimpft *[den Politiker]}
who$_{ACC}$ the$_{NOM}$ journalist$_{NOM}$ attacks *[the$_{ACC}$ politician]

d. den der Journalist beschimpft *[der Politiker]
who$_{ACC}$ the$_{NOM}$ journalist$_{NOM}$ attacks *[the$_{NOM}$ politician]

..., liefert die Beweise.
..., delivers the evidence.

Participants were presented with the sentences word by word and had to click a button as soon as they found an error. Crucially, although all target sentences are ungrammatical, the incorrect continuations differ with respect to their local expectability. While the nominative NP in SRCs completes a locally coherent (topicalized) main clause structure (NP_{ACC}-V-NP_{NOM}, (22b)), the accusative continuation does not constitute local coherence (22a). In ORCs, the accusative NP completes a canonical NP_{NOM}-V-NP_{ACC} sequence (22c), but not the nominative NP (22d). The rationale underlying the experiment is that if participants built up expectations based on local context, they should exhibit longer error detection times if an upcoming element matches this locally induced expectation, because the locally expected element will be preactivated and thus be harder to reject. Konieczny (2005) indeed found elevated error detection times for locally coherent continuations, however, only in SRCs. For ORCs, only participants with high levels of reading-experience showed a tendency in the expected direction. Although the pattern of effects is somewhat inconclusive, the results provide evidence that local structure can induce the prediction or, at least, expectation/preactivation of elements that are not licensed by the global syntactic structure of the input sequence, and thus resemble the observed false alarm activations in SRNs following locally coherent sequences.

2.3.3 No Need for Local Coherence? Word-Based Accounts

2.3.3.1 Lexical Bottom-up Interference

Gibson (2006) was the first to attempt to explain the results from Tabor et al. (2004) without the strong commitment to self-organized processing put forward by Tabor and Hutchins (2004) and Tabor et al. (2004). Gibson's approach mainly aims at the dynamical *Visitation Set Gravitation* (Tabor et al., 1997; Tabor & Tanenhaus, 1999), and the empirical results that have been interpreted as evidence for the sensitivity to syntactic context in resolving lexical ambiguity. Gibson's proposal assumes, similarly, that the system is sensitive to syntactic and lexical expectability at a particular word in a given syntactic context, representing a top-down predictive component. In addition, the system is also assumed to be sensitive to relative frequency of lexical category distribution of the actual processed word. Crucially, the latter component is conceived as *context-independent*, such that only global top-down expectations, on the one hand, and bottom-up, word-based information, on the other hand, are conceived as interacting sources of processing complexity and, hence, predictors of reading time. Global predictability is formalized via the relative frequency of part-of-speech tags given the word in the particular syntactic context. Crucially, an initial value of 0.1 is added to the derived syntactic expectability, such that also ungrammatical part-of-speech tags receive a value above zero. The context-independent bottom-up component is operationalized as the relative frequency of part-of-speech tags given the word, but irrespective of context. Both values are then multiplied and normed with respect to other occurring categories of the actual word, resulting in a normalized Lexical bias LB_{norm}. Assuming, for example, a probabilistic, serial parser, the resulting value represents the probability for each part-of-speech tag to be assigned to the actual word, resulting in immediate reanalysis if the chosen one can not be integrated with the syntactic structure, thus leading to elevated processing time for lexically ambiguous words. Here, I will only consider Gibson's account to explain the results from Tabor et al. (2004), and not discuss the series of experiments conducted to provide evidence for the proposal in general.

Based on corpus counts, Gibson observes a higher LB_{norm} for the (locally coherent) main-verb reading of the ambiguous participles (eg., "tossed"), compared to the unambiguous one (e.g., "thrown"), due to their different lexical category probabilities. While the difference between ambiguous and unambiguous elements is explained by the context-independent bottom-up component, the missing difference in the unreduced cases is explained by differences in top-down predictability due to the unreduced relative clause that more strongly biases the globally correct past-participle reading.

Gibson argues, on the one hand, by due of parsimony of the model, since only global syntactic expectations and relative category frequency have to be tracked and stored, but not category frequencies relative to the local context, while in local coherence accounts the parser also has to keep track of and store relative category frequencies in various local syntactic environments. On the other hand, Gibson claims that his proposal can account for a variety of results regarding lexical category disambiguation which do not involve local syntactic coherence of any sort.

A crucial component of the proposed measure is the *smoothing* of global predictions that assign an above zero probability even to ungrammatical tags, which is responsible for the fact that context-independent lexical bias of grammatically unlicensed tags can have any influence at all, as in cases of local syntactic coherence. The particular smoothing value as well as the procedure of smoothing in general can be conceived as a potential weakness of the approach (see also Bicknell et al., 2009, for this point). In addition, the approach can not readily account for experiment 2 in Tabor et al. (2004), although Gibson give a tentative description of how the model could be extended with regard to semantic plausibility Gibson (2006, p. 383). Gibson, however, does not make any strong claims about which mechanisms may give rise to the influence of lexical bias and its components, although he proposes that both a serial processing system – by due of stochastic processes – as well as parallel approaches assuming competition between lexical entries are compatible with his proposal.

To conclude, Gibson provides a model – or rather, a measure – that falls under the class of self-consistent parsing accounts. The proposed interaction between top-down predictions and a word-based bottom-up component can, at least partially, account for the results reported in Tabor et al. (2004), and will thus have to be considered as an alternative to the self-organized parsing approach. However, only relying on lexical category ambiguity, the approach would hardly explain similar effects where no ambiguity is involved.

2.3.3.2 Retrieval Interference in Cue-Based Parsing

The influence of grammatically unavailable constituents on processing has also been investigated within the cue-based parsing approach described earlier (section 2.1.2.2). Within this framework, constituents have to be retrieved from memory to be integrated with their head, while retrieval is accomplished by retrieval cues that specify features of the elements to be retrieved. Crucially, the retrieval process is subject to interference if more than one recently processed element satisfies these cues. Inherent in this proposal is that constituents that are already syntactically integrated and should therefore not be available anymore, can be retrieved and incorrectly attached if they happen to be retrieved instead of the correct constituent. Thus, as local coherence accounts claim, the already processed input does not fully inhibit "ungrammatical" attachments. However, the results are explained by referring to different mechanisms, and thus are not understood as effects of local syntactic coherence: The factor that determines whether retrieval interference takes place, and how strong these interference effects while processing the head will be, depend solely on the feature-overlap of preceding NPs, but not on the local syntactic context. Thus, as in Gibson (2006), the effects are not taken to reflect partial, grammatically impossible subparses that interfere with the global sentence structure. In a series of experiments, Van Dyke (2007) tested if semantic and syntactic feature-overlap between constituents can be shown to be a source of processing complexity. Semantic similarity was achieved by nouns that were semantically plausible or implausible arguments of the verb ("the neighbor/the warehouse ... was complaining"). Syntactic similarity was operationalized as subject- (*HighSyn*) or objecthood (*LowSyn*) of the intervening noun (23).

(23) The worker was surprised that the resident

a. who was living near the dangerous warehouse was complaining about the investigation

b. who was living near {the dangerous neighbor was complaining about the investigation}

c. who said that the warehouse was dangerous was complaining about the investigation

d. who said that the neighbor was dangerous was complaining about the investigation

(Van Dyke, 2007, example stimuli from experiments 1 and 2)

In a self-paced reading experiment with a "Got it" task, Van Dyke (2007) found both main effects for semantic and syntactic interference of the "Got it" response, and a main-effect for semantic interference for reading-times on the critical region "was complaining". However, the results are, as the author admits, mainly compatible with local coherence accounts, since the clearest effect for reading times was one of semantic similarity in the *LoSyn* conditions, where reading times where higher for example (23b) – which is the condition that contains the locally coherent sequence "the dangerous neighbor was complaining". Since the factor of syntactic interference is realized by an additional subordinate clause, it is at least questionable if the observed main effect of syntactic interference in the "Got it" task (higher rate of "yes" answers for *LoSyn*) has to be attributed to interference. The claim that a main effect of semantic interference (and interaction) on the last word provides evidence for semantic interference is, given that it arises rather late, at least debatable. The observed results in an additional eye-tracking while reading experiment using the same materials also does not conclusively rule out local coherence accounts. Finally, in a third experiment Van Dyke (2007) report additional evidence for the influence of retrieval based interference. However, since the design does not include local coherent sequences, I will not discuss it here.

In general, the results reported in Van Dyke (2007) provide, in spite of some rather inconclusive aspects of the data, evidence that feature overlap between constituents can lead to retrieval of grammatically unavailable elements, which stand in stark contrast to the self-consistency assumption underlying grammar-oriented approaches to human sentence processing. In this respect, the cue-based retrieval approach corresponds with the claim of Tabor et al. (2004), but ascribes the effects to different mechanisms. Crucially, while the claim that local syntactic coherence may induce processing difficulty implies that the local sequential context plays a role, cue-base retrieval accounts ascribe the effects to retrieval errors that are exclusively related to the noun that is attempted to be integrated. In this regard, the model also differs from Gibson's (2006) account, where the difficulty is explained as arising from lexical category ambiguity.

Since the materials used in Van Dyke (2007) do not include any ambiguous elements, Gibsons approach can not readily account for the data. In my view, Van Dyke's (2007) results can not readily rule out influences of local syntactic coherence. In addition, the cue-based parsing approach can hardly account for the results from experiment 1 in Tabor et al. (2004), since the preceding context of the ambiguous element is identical in all conditions, so that no differences in interference and the retrieval dynamics are to be expected. With regard to experiment 2 in Tabor et al. (2004), stronger interference should be expected when subject (*bandit*) and object (*prisoner/gold*) share animacy, resulting in a main-effect of animacy, which, however, was not observed. In general, since the approach mainly makes predictions about integration processes at the verb, it is difficult to clearly evaluate the proposal with regard to other effects (as, for example, the anomaly detection results from Konieczny, 2005). In general, evidence for retrieval interference has been provided independently in a number of experiments. Thus, it may well be that both local syntactic context and retrieval mechanisms may independently influence processing. The current results so far, however, do not allow us to make any conclusive claims in this respect.

2.3.3.3 Preliminary Discussion

Both Gibson's (2006) and van Dyke's (2007) proposals explain the observed effects not as stemming from local syntactic coherence – a globally incoherent subparse interacts/interferes with the globally warranted structural analysis – but try to account for the effects by positing interference between the parsing mechanism and context-independent lexical retrieval processes[15]. In Gibson's (2006) account, retrieval of the (lexically ambiguous) verb is proposed to determine processing difficulty, whereas cue-based parsing ascribes the observed effects to properties of the constituents that have to be retrieved when the verb is encountered. Depending on feature overlap of recently processed nouns in the sentence, this can result in retrieval interference. Both proposals can partly provide alternative explanations for results associated with local syntactic coherence, but can, as far as I see, neither explain all observed patterns, nor rule out the building up of partial, locally induced subparses as a source of processing complexity.

[15] Gibson does not explicitly posit a processing model, but in our view the proposal can well be explained in the framework of lexical retrieval.

2.3.4 Modular Approaches

2.3.4.1 A Module for Lexical Category Disambiguation

Another class of models predicts effects of local syntactic coherence, at least for cases where it is induced by lexical category ambiguity, as in Tabor et al.s' results. This class of models, referred to as *Local Coherence Accounts* 2 (LCA2) in Tabor et al. (2004), proposes a process or module that operates on a limited window-size that, in contrast to Gibson's approach, spans more than just one word. The results of this module are passed to the next processing step, which typically operates on a sentence-wide basis. The sausage-machine (Frazier & Fodor, 1978) described in section 2.1.1.2 would fall under this class. More recent approaches in this line are, for example, the modular statistical architecture proposed in Corley and Crocker (2000); Crocker and Corley (2002); Crocker and Brants (2000), and the information-theoretically based approach by Bicknell et al. (2009). In contrast to the sausage machine, both proposals assume that acquired knowledge about distributional properties of the language are the main source of information the module is operating on.

Corley and Crocker (2000) proposed a model of human sentence processing that is based on incremental cascaded markov models (see also Crocker & Brants, 2000). The model implements the modular statistical hypothesis, stating that different steps in the process of understanding a sentence are accomplished by several modules that operate on statistical information within their respective structural level. A crucial ingredient of the model is a module for lexical category disambiguation (basically a POS-tagger), operating on unigram statistics of a word (similar to Gibson, 2006) and on POS-based bigram[16] transitional probabilities, amounting to (2.2):

$$p(POS_{w_n}|w) * p(POS_{w_n}|POS_{w_{n-1}}) \tag{2.2}$$

The best hypotheses about the part-of-speech of a word are then passed to the next (syntactic) module. If the POS can not be integrated in the already built up structure, reanalysis has to occur. Corley and Crocker (2000) review two self-paced reading experiments involving lexically ambiguous words (verb/noun) as evidence for their proposal, showing that statistical bias of the ambiguous elements interacts with the interpretation required by the following disambiguating region. This provides evidence for the influence of statistical information on initial category decision. In a second experiment, the statistical bias of the ambiguous elements also lead to elevated reading times, although the biased interpretation was not compatible

[16] In Crocker and Brants (2000), the module also incorporates trigram information.

with the preceding syntactic context. The latter results, in particular, are taken as evidence for a pre-syntactic module that achieves initial decisions about lexical category before and independently of proper syntactic analysis.

Thus, Corley and Crocker (2000) assume bottom-up processes with a limited context-dependency (one or two preceding words). Although local syntactic coherence is not explicitly considered or modeled in the proposal, the lexical category disambiguation module should, e.g. for Tabor et al. (2004)'s stimuli, pass the locally coherent POS as the best hypothesis to the next layer, which will in turn result in reanalysis and elevated processing effort.

Corley and Crocker (2000) explicitly propose a strictly modular architecture, both on account of computational efficiency and methodological grounds. In particular, the subscription to a modular architecture is aiming at interactive constraint-based accounts that are accused of being hard to evaluate with regard to broad coverage of different linguistic phenomena. The extensive use of statistical information, on the other hand, contrasts with earlier, purely structure-based accounts such as the Garden-Path model as well as the sausage-machine. With regard to (some instances of) interactive constraint based approaches, Crocker and Corley (2002, p. 20) point out that any model proposing that initial lexical category decisions are determined by the syntax do not predict effects of lexical category ambiguity in cases where syntactic constraints provide information about the correct interpretation – which is exactly what local syntactic coherence amounts to. The authors argue that "some possible interactive constraint-based models" (Crocker & Corley, 2002, p. 22) fall under this class. This basically amounts to the claim that interactive constraint-based accounts that imply strictly self-consistent parsing will not predict influence of local syntactic coherence (or, more general, influence of local context). As I pointed out earlier, it is not entirely clear if or which interactive constraint based accounts subscribe to this principle. There may well be constraints operating on local context that take part in the constraint satisfaction process and would thus allow both influence of global syntactic information on disambiguation as well as influence of local context. Crocker and Corley (2002, p. 24), however, argue that, although such constraints could be stipulated (as well as constraints producing the opposite effect), the modular approach provides a methodologically preferable model where the prediction of local, bottom-up influences are directly predicted through the very architecture of the system, and do not have to be stipulated to account for existing data.

The modular statistical hypothesis is thus compatible with effects of local coherence, at least with regard to experiment 1 in Tabor et al. (2004). The interaction observed in experiment 2 is, as Tabor et al. (2004, p. 362) note, not predicted if only rather coarse grained category information about the preceding word is taken into account, which has been claimed to be the case at least with regard to number information Crocker and Corley (2002, p. 22). However, reconsidering the granularity of information available for the pre-syntactic module could extend the model to also account for the results of experiment 2 in Tabor et al. (2004).

To conclude, the approach includes both statistical information on the word-level, comparable to Gibson's approach. The crucial difference, however, is that information used for lexical category disambiguation is not context-independent, but also takes into account a limited amount of preceding context. The same difference holds true with regard to the cue-based parsing approach, which attempts to account for difficulty without referring to the local sequence, but with regard to noun phrases that have been encountered before.

2.3.4.2 Local Beliefs

A rather similar approach, at least with regard to the statistical information used for lexical category disambiguation, has been proposed by Bicknell et al. (2009). First of all, they provided further evidence for effects of local coherence in an analysis of the Dundee reading corpus (Kennedy & Pynte, 2005), showing that effects of local syntactic coherence do not only occur in manually constructed, isolated sentences, and, referring to the stimuli used in experiment 1 in Tabor et al. (2004), rather rare structures, but also in reading naturalistic, coherent texts. Local coherence was identified as "a string of words w that out of context would suggest one very likely parse, and that parse is impossible (or at least highly unlikely) in context.". Bicknell et al. (2009) included one-word local coherence (where w is a string of length 1), assessed as the probability of a part-of-speech tag t given w , $p(t_i|w_i)$, as well as two-word local coherence, measured as $p(t_i|w_{i-1}^i)$. Reading times where found to be higher for cases where the locally expected tag differed substantially from the globally correct tag, while the effect was even stronger for two-word local coherence.

While the effects of one-word LSCs are compatible with both Gibson's (2006) context-independent approach as well as the dynamical-system approach in Tabor et al. (2004), the stronger effect of two-word LSCs can only be accounted for the latter model. Bicknell et al. (2009), however, argue that the SOPARSE model (Tabor & Hutchins, 2004) implies a "rather arbitrary parsing mechanism with a large number of free parameters" (Bicknell et al., 2009, p. 11). The authors propose an alternative model that they claim overcomes these problems.

As in Levy (2008a), the basic theoretical framework of the model is Bayesian, construing the process of incremental comprehension as the update of beliefs about the likely structure of a word sequence carried out with every incoming element. In contrast to a purely surprisal-based model, however, the Bicknell et al. (2009) model also incorporates guesses about the next element that are derived from a local (n-gram) string sequence, and thus allows for bottom-up effects that are independent from the global context. N-gram based predictions result in a *prior* distribution representing the likelihood of the POS-sequence associated with the partial string. This prior distribution is compared with the posterior distribution over possible structures given the sentence-wide syntactic requirements at the currently processed word in the input string.

Processing difficulty is formalized as the informational difference between prior and posterior distribution, representing the amount of change that is necessary to update the beliefs about the interpretation of a sequence from the locally-derived *prior* distribution over possible tag-sequences, to arrive at the *posterior* distribution informed by the entire context of the sentence processed so far.[17]

The overall outcome of the model is an incremental word-based measure of processing-complexity. Thus, the authors are able to model effects of local coherence, since the presence of locally coherent sequence results in a prior distribution that requires a substantial change to arrive at the posterior, representing the distribution of possible globally consistent parses.

Bicknell et al. (2009) argue that their approach can be taken as a extension of Gibson's (2006) approach. However, the central claim of Gibson's approach is that the effects can be explained by the interaction of top-down predictions with context-*independent* unigram-statistics alone, leading to a more parsimoniuos model. Thus, the "extension" would basically discard the central claim of the Gibson's (2006) approach.

Although framed in an information theoretical framework and referring to belief-update, the approach is, in relevant aspects, closely related to the *modular statistical hypothesis* proposed in Corley and Crocker (2000). In contrast, however, Bicknell et al. (2009) provide a rather abstract formal account, and do not make any strong claims about the architectural implications of this formalism. Rather than providing an implemented (processing) model of as what sentence processing could amount to, it merely provides a rather abstract functional description of behavioral data, but no explicit proposal of the cognitive processes or mechanisms that could instantiate the formalism. The authors, however, claim that

> This belief update process can be viewed as a rational reconstruction of the SOPARSE model by Tabor and Hutchins (2004), where — instead of the system dynamics

[17] A detailed description of the underlying belief-update process can be found in Bicknell and Levy (2009).

> of competition between arbitrary tree fragments -– differences between prior and posterior probability distributions over syntactic structures determine processing difficulty. (Bicknell et al., 2009, pp. 9)

However, it is not entirely clear what such a "rational reconstruction" amounts to. In particular, it is not clear if the authors claim that their proposal can be understood as a description of what the dynamical system does – in other words, specify the function the system computes – or if they propose alternative *mechanisms*. If the latter, these mechanisms are not explicitly articulated. If the former, the claim basically boils down to specifying why the system, under the assumption that it is optimally adapted to its task – will inevitably produce phenomena of local coherence as a side effect of an overall superior mechanism. As Bicknell and Levy (2009, p. 672) state, calculating the prior distributions in the first place is only an advantage if the priors have been calculated in advance and can be accessed relatively quickly during processing. Thus conceived, however, the proposal at least to implies that information on part-of-speech co-occurrences and statistical information is acquired, stored and retrieved, and therefore seems to imply at least some module that does this task.

2.3.4.3 Preliminary Discussion

To conclude, Bicknell et al. (2009) provide further evidence for the influence of local syntactic coherence in reading, and provide an expectation-based formalism that is able to model the data. However, the level of description of the model is not fully specified. The general argument that both the accounts by Corley and Crocker (2000) and Bicknell et al. (2009) imply is that bottom-up information drawing on statistical properties of local substrings of the input is used as a valuable source of information to achieve efficient processing. In cases, however, where the local information conflicts with global syntactic requirements, this mechanism can be a source of processing difficulty.

The assumption of an *n-gram* based module or process is explicitly considered under a perspective of efficiency, or, in other words, as a rational strategy given the requirements and bounds of the processing system. The same holds true for Gibson's (2006) approach, where taking into account only context-independent relative category frequencies is proposed as a rather "cheap" way of making advantage of the information held within statistical properties of the language.

In contrast, the question of efficiency or rationality is at the core of neither Tabor et al.s' (2004) nor and Van Dyke's (2007) models, which instead argue on the grounds of general principles of cognition that lead to the effects. Put simply, neither retrieval interference nor interference of ungrammatical subparses are explicitly construed as unwarranted side-effects of an otherwise particularly efficient strategy, but rather as side-effects of the general principles and mechanisms the system is based on. However, of course, these general mechanisms still have to provide the capability of performing the task to a sufficient degree. The question of how to incorporate the reported results within a framework of rational parsing is more explicitly stated in a further group of attempts that I will now describe.

2.3.5 How to Integrate Local Coherence in Rational Parsing?

2.3.5.1 A Rational Parser Would Do It: Parsing as Informed Search

In 2011, Hale published a paper with the speaking title: “What a rational parser would do”, that, among a variety of other phenomena, explicitly attempts to incorporate effects of local syntactic coherence into a theory of parsing as a rational enterprise. The basic idea underlying the approach is that parsing can be understood as an informed search process. The search space is defined by the possible syntactic structures of a language, and the incremental input guides this search to finally end up with a consistent interpretation. Each incoming word or element constrains the search space insofar as it provides information about the likelihood of which syntactic structure will finally map onto the input sequence.

The search process is implemented by adopting the A* search algorithm (Hart, Nilsson, & Raphael, 1968). It incorporates, on the one hand, the costs – in parser steps[18] – that have been taken so far to reach a particular node *n*. On the other hand, completion costs are taken into account as the expected number of steps or intermediate states to reach a completed parse starting from the actual node *n*. The number of expected steps to completion is estimated as the average number of steps that were necessary to finish parsing in similar situations, based on corpus probabilities. These two measures are balanced to arrive at a reasonable solution without with reasonable effort, or, in Hales words, “without wasting too much time exploring unpromising subspaces” (Hale, 2011, p. 407). Using simplified grammars and corpus-based distance-to-completion measures, Hale shows that his model is able to both account for garden-path effects and common counterexamples of purely structural preferences. Crucially, the model also passes more intermediate steps in sentences that contain locally coherent sequences, as in experiment 1 in Tabor et al. (2004). Due to the rather high expected comple-

[18] Steps are estimated by a parser implementing generalized left corner parsing.

tion costs in the locally coherent case, more intermediate steps are visited before the parser faces eminent costs for the wrong path, forcing the mechanism back onto the correct path.[19] Hale also provides a simulation of the German sentences from Konieczny, Müller, Hachmann, Schwarzkopf, and Wolfer (2009)[20], showing that more intermediate states are visited when a local syntactic coherence is present.

Hale explicitly attempts to provide a *rational* approach of parsing and argues for a symbolic parsing mechanism as an alternative to the dynamical system approaches put forward by Tabor et al. (2004) and Konieczny (2005, ...). However, it seems clear that rationality in this case can not be easily identified with self-consistency anymore, since, in contrast to other symbolic parsing models, no mechanism explicitly prevents the parser from exploring ungrammatical, unfinishable states at a particular step (Hale, 2011, p. 421). It is, however, not entirely clear to me if the model can account for local coherence effects that do not involve lexically ambiguous items, since the model is rather hard to evaluate without actually running it. However, this remark is not meant to be a criticism of the approach, but rather owing to the limited resources of time and space within this work.

2.3.5.2 Reconsidering the Input: the Noisy-Channel Approach

In Hale (2011), the crucial requirements that inform what would be "rational" for the parser have been the task to quickly and accurately arrive at a structural interpretation of the input string, leading to a search strategy that provides a good balance between these requirements, given the available resources and information. Another path of considering effects of local syntactic coherence as resulting from a rational system has been taken by Levy (2008b). In contrast to the majority of other approaches, the focused bounds on rationality with regard to language comprehension in this approach are not merely speed, accuracy and limited resources, but rather the nature of the signal itself. A rational processing system not only has to make efficient use of its resources, but also has to take into consideration that there may have been errors in the signal. Such errors can either stem from the speaker, from noise disturbing the transmission, or from perception itself.

For example, if a comprehender encounters an element that is highly unexpected in the given context, she will not only have to update the hypothesis about the structure that could

[19] Although Hale (2011) reports more intermediate steps for the ambiguous compared to the unambiguous states, it is not entirely clear from the paper in which way the relative clause type (reduced/unreduced) in the design of experiment 1 in Tabor et al. (2004) influences the number of intermediate steps.

[20] We will give a detailed description of the data in section 3.3.

possibly be assigned to the input, but will also consider that there is a chance that an element of the input has been misunderstood, incorrect or missing. Considering this error may, in some cases, be more plausible and, hence, more efficient than trying hard to find a consistent interpretation assuming a perfect input signal. Thus, the proposal assumes that during incremental processing of a sentence the beliefs about the likely interpretation of a string are not only influenced by predictability of the word that is actually processed, but also by estimations of the likelihood of errors in the already processed input in the face of the actual input word.

The rationale with respect to the results from Tabor et al. (2004) is that a slight change in the stimuli would render the local interpretation globally correct, for example by adding a word

(24) The coach smiled at **how** the player tossed the frisbee.

or by changing some letters.

(25) The coach smiled a**nd** the player tossed the frisbee.

In such cases, it could be very likely that in the face of the rare and complex construction we would think that there was or we made an error beforehand rather than trying hard to make sense of the sentence as it is. Without going into the details of the calculus underlying Levy's (2008b) proposal, it should be mentioned that the incorporation of potential errors in the linguistic signal or channel is a quite compelling approach when considering rationality given a noisy environment. Levy, Bicknell, Slattery, and Rayner (2009) provide further evidence for the approach by conducting an eyetracking-while reading study. They used the sentences from experiment 1 in Tabor et al. (2004), but manipulated the neighborhood of the element before the local syntactically coherent sequence. For example, they substituted the *at* in (20a) and (20b) by *toward*, which does not have near-neighbor words under which *the player* would be a grammatical subject.

(26) The coach smiled toward the player tossed the frisbee.

(27) The coach smiled toward the player thrown the frisbee.

Thus, local syntactic coherence remains the same as in examples (20a) and (20b), and the SOPARSE model would thus predict the same difference between examples (26) and (27). However, Levy et al. (2009) replicate the difference between (20a) and (20b), but found no difference in a variety of measures when using "toward". However, in first-pass reading-times, they found no significant interaction of *Preposition* $\times$ *Ambiguity*, but a significant main-effect of *Ambiguity*. Thus, the results are somewhat inconclusive. On the one hand, the number of regressions out of the critical region, question accuracy and go-past times seem to support the noisy input model, while first-pass times are compatible with the explanation given in Tabor et al. (2004). One interpretation would obviously be that both processes – local coherence processing and uncertainty about preceding input – are at work. The missing interaction in first-pass times can also be interpreted as a purely lexical effect, since the conditions with unreduced relative clauses were not included in the experiment. In my view, the approach does not fully explain the results, but at least has to be considered as an interesting and powerful hypothesis. The approach has recently been taken up in Piantadosi et al. (2012). As has been pointed out with regard to Bicknell et al. (2009), however, the approach also does not propose an actual processing model in the strict sense, but rather aims at giving a functional description of what the human language processor does and why it does what it does. It does, however, not tell us much about how this is achieved, apart from the fact that statistical properties play an important role within the mechanism.

2.3.5.3 Local Coherence as a "Good Enough" Parse

Similar to Levy's (2008b) focus on conditions of communication, another strand of research has emphasized the importance of the circumstances of "real-world" language comprehension. Rather then focusing on the requirements posed through a noisy input channel, it emphasizes the idea that sometimes it simply may not be necessary to completely understand what has been said in a conversation. The *good enough parsing* approach (Ferreira, 2003; Ferreira & Patson, 2007) is questioning the assumption often underlying comprehension models that the outcome, or even the goal, of comprehension is always a complete, consistent and accurate representation of the meaning intended by the speaker. Instead, it is proposed that in many situations, fast and frugal heuristics are sufficient – or *satisficing* – for the goals at stake, so to as roughly follow the conversational topics. In general, the approach can explain the effects of local syntactic coherence quite well as the result of fast and frugal heuristics operating on surface probabilities. Given that the embedded sequence mirrors a rather regular and frequent sentential structure, a shallow, heuristic parse may grab the local interpretation without much effort, and either ignore the global structure totally or at least maintain the local interpretation in parallel to some degree even after noticing the globally correct interpretation.

The approach can be seen as a more recent consideration of the perceptual strategies proposed in Bever (1970) and incorporates an aspect of communication that has been mostly ignored in theories of sentence processing. Since the theory is rarely implemented as a model[21], it is not easily evaluated with regard to the different results reported so far. It implies a system that can be said to be rational with regard to the task, but does not subscribe to the self-consistency assumption. In general, it is compatible with the effects on processing time and evidence for interpretational processes triggered by local syntactic coherence, which I will describe in the following section.

2.3.6 From Processing Load to Interpretation: Evidence from the Visual World Paradigm

Both the results and the theories described so far mainly regard reading, measuring and explaining differences in reading time between sentences that contain local syntactic coherence or not. However, given that local coherence does influence processing, the question arises about how deep these ungrammatical sequences are processed. Are locally coherent sequences only processed at some intermediate level, resulting in disturbance, additional processing load and slowed down reading? This would be incorporated in any model that includes a selection process of some sort which, after having processed the sequence, rules out the ungrammatical subparse, and ensures that only the globally consistent parse is passed on and gets interpreted (or, if more than one is possible, as in cases of structural ambiguity, possibly a number of them). Such an approach could be still said to subscribe to the notion of self-consistency formulated in Frazier and Clifton Jr. (1996, p. 3), that "[...] readers and listeners do no violate their knowledge of grammar in arriving at an interpretation of a sentence" and that "They do not arrive at interpretations that violate grammar.".

Experiment 2 in Tabor et al. (2004), however, already provided some evidence that not only structural, syntactic information (e.g., part-of-speech information) determines if local coherence influences processing, but that semantic coherence within the local sequence – operationalized by manipulating animacy of the "local subject" – plays a role. The influence of semantic properties such as animacy is incorporated in the SOPARSE model by the fact that, in contrast to Vosse and Kempen (2000), both syntactic and semantic features contribute to

[21] An attempt to explicitly model at least some of the results from Tabor et al. (2004) based on "good enough parsing" is reported in Morgan, Keller, and Steedman (2010). The authors present a strictly bottom-up parser that takes into account the ratio of inside probability of subsequences and sentence probability and thus can model effects of local sequences that are locally very likely to have an interpretation that is not consistent with the global context. The model, however, only focuses on the data from grammaticality judgements in Tabor et al. (2004).

the linking dynamics of the system. In addition, Tabor et al. (2004) ran a follow-up study (pp. 363) where participants were asked questions referring to to the global or the local content. The proportion of positive answers for agent-related questions ("Did the prisoner transport something?") was significantly higher for the local coherence conditions. This provides evidence that the meaning of the local subsequence is, at least sometimes, extracted and available. However, the effects are rather small. In addition, the slightly higher proportion of positive answers may be a result of reconsidering the sentence in the face of the question, resulting in a rough match that may suffice to answer "yes" in some cases. Thus, the result may not reflect processes that happened during reading, but rather mirror later processes related to the particular task [22]. A different way of testing whether local syntactic coherence does trigger interpretation of the local sequence has been the visual world paradigm. The rationale is that if the local content is being interpreted, it should trigger eye movements to a scene depicting aspects of the local content, even if this scene is not compatible with the global content of the sentence.

Evidence for the interpretation of locally coherent sequences has been provided in two visual world experiments using complement clauses with embedded locally syntactic sequences (Konieczny & Müller, 2006; Konieczny, Müller, Hachmann, et al., 2009) and dative relative clauses (Konieczny, Müller, Baumann, et al., 2009). In both experiments, participants were shown to be more likely to fixate a scene depicting the local content when listening to sentences with local coherence than when listening to control sentences without. Both experiments will be described in detail in chapter 3.

2.3.6.1 Influence of Local Coherence on Anaphora Resolution and Binding Constraints

Weldle (2011) conducted a series of visual-world experiments to test if local coherence can influence anaphora resolution. Participants where presented with sentences with reflexives or pronouns, where a locally coherent sequence induced different binding than the global sentence (28).

(28) Als der Vater$_i$, den …
While the father$_i$, that …
"While the father that …

a. {der Sohn$_j$ kämmt, sich$_{i/*j}$ im Wohnzimmer} anzieht,
the son$_j$ combs, himself in the living room dresses,

[22] See also Tabor et al. (2004, p. 367) for this point

the son combs dresses himself in the living room,

b. {der Sohn$_j$ kämmt, ihn$_{*i/j...}$ im Wohnzimmer} anzieht,
the son$_j$ combs, him in the living room dresses,
the son combs dresses him in the living room,

c. der Sohn$_j$ gerade kämmt, sich$_{i/*j}$ im Wohnzimmer anzieht,
the son$_j$ just combs, himself in the living room dresses,
the son just combs dresses himself in the living room,

d. der Sohn$_j$ gerade kämmt, ihn$_{*i/j...}$ im Wohnzimmer anzieht,
the son$_j$ just combs, him in the living room dresses,
the son just combs dresses him in the living room,

...ist die halbe Stadt bereits auf dem Weg zur Aufführung.
...is the half town already on the way to the show.
...half the town is already on their way to the show."

Considering example (28a), globally the reflexive «sich» [himself] refers to the father who is combing himself. However, when considering only the local sequence «der Sohn kämmt (,) sich», the reflexive refers to the son combing himself. Respectively, in example (28b), the pronoun «ihn» [him] globally refers to the son (being dressed by the father), while in the local context the pronoun refers to the father (being combed by the son). While listening to the sentences, participants where simultaneously presented with different visual scenes depicting local and global content (as well as distractor scenes). For both reflexives and pronouns, the results revealed higher fixation proportions on the scene with local coherence when participants listened to a sentence with local coherence than when listening to control sentences without local coherence (examples (28c) and (28d), where local coherence is inhibited by a inserted «gerade» [just], leading to an ungrammatical subsequence). The results provide evidence that participants are sensitive to the local coherent sequence and also partially process anaphoric references within this local context which can conflict with global binding constraints.[23]

Weldle (2011) developed a connectionist model of anaphora resolution that combines a simple recurrent network with an additional referential layer where each output-node represents one of the referents. The network was trained on predicting the next word of the sentence as well as on activating possible referents at the referential layer when presented with a pronoun or reflexive in the input sequence. In case of local coherence, the networks activated, in addi-

[23] For further, similar experiments see Weldle (2011), chapter 7.

tion to the globally licensed referent, also referents that fall within the 'local' binding domain, mirroring the results observed in the visual-world experiments.

2.3.6.2 Going for the Local Arguments

Further evidence for the interpretation of local syntactic coherence has been provided in a visual-world experiment (Konieczny, Müller, Baumann, et al., 2009) using sentences without lexically ambiguous elements (29). In these sentences, the locally coherent sequence is constructed by using object-extracted relative clauses, where the verb phrase of the relative clause «die Floristin beschenkte» [the florist gave a present to] and, crossing the RC boundary, the object of the main clause «den Müllmann» [the garbageman] constitutes a canonical main clause structure, representing a different proposition than the main clause. Participants showed a higher likelihood to inspect a scene depicting the local content, but only when the strong prosodic cues marking the clause boundary where eliminated by splicing in the local sequence. An effect even showed up for (29b), where the perfect-tense induces a weaker local coherence as the beginning of a topicalized main clause «beschenkt hat den Müllmann . . . ». Although missing the subject, this sentence fragment, as the past-tense sequence in (29a), represents the garbageman as the recipient of the RC verb, in contrast the the globally correct recipient, the postman.

(29) Als der Postbote, den . . .
When the postman, that . . .
"When the postman that . . .

a. {die Floristin beschenkte, den Müllmann auf der Straße}
{the florist presented [with something], the garbageman in the street}
the florist gave a present to met the garbageman in the street

b. die Floristin {beschenkt hat, den Müllmann auf der Straße}
the florist {presented [with something] has, the garbageman in the street}
the florist has been giving a present to met the garbageman in the street

. . . traf, wurde es gerade dunkel.
. . . met, got it just dark.
. . . *it was just getting dark."*

The visual world results provide evidence that local syntactic coherence can not only provide a source of processing complexity, which could be accounted for within a modular approach that generates a temporarily active, globally incoherent subparse, based on fast and frugal heuristics or a statistically driven preprocessing step operating on a limited window size of several words. In contrast, these local substructures do not seem to be totally eliminated in later steps of processing, but can also bring to bear actions, for example eye-movements, that operate on the meaning of the local sequence. This aspect of immediate action based on bottom-up information is considered explicitly in Kukona and Tabor (2011), who also provide evidence for effects of local, bottom-up information that can lead to, in some sense, "irrational" eye movements in a series of visual world experiments. They also provide a dynamical system model implementing their *impulse-processing* approach, stating that eye movements, and, in general, action, are driven by continuous impulses from several perceptual sources as is, with regard to the visual world paradigm, language and visual information.

Since the experiments do not, in the strict sense, show effects of local *syntactic* coherence, we will not report the results here, but will refer to the underlying model later on.

2.4 Summary and Discussion

At the beginning of the chapter, I reviewed two strands of research that have been highly influential in psycholinguistic research. With regard to ambiguity resolution, a main topic has been, on the one hand, the question whether the human sentence processing mechanism is better conceived as a serial or parallel mechanism, and, consequently, if processes of reanalysis or competition are responsible for the processing difficulties that arise – at least in some cases – of ambiguity. The sources of information that are used to resolve ambiguity have been mainly at stake with regard to the question if the processing system has to be conceived as a modular or an interactive device. Research on ambiguity resolution has established the insight that sentence processing is incremental, trying to integrate each incoming item immediately with the preceding input and has led to the closely related assumption of self-consistent parsing.

Linked with these questions, and often associated with particular assumptions regarding them, the proposed mechanisms either focused on the role of a limited working memory as the driving force that determines which strategies are used to resolve ambiguity, or emphasized the role of experience and, consequently, mechanisms of prediction as crucial properties of the system.

The opposition between resource-based and experience-based accounts is also vivid with regard to research on online processing difficulty. I argued that effects of local syntactic coherence provide a phenomenon that strongly supports, on the one hand, experience-based approaches and the crucial role of prediction in sentence processing. On the other hand, effects of local syntactic coherence are not compatible with some recent prediction-based measures of online processing difficulty that explicitly emphasize the *rationality* of the human sentence processing mechanism. In contrast, effects of local syntactic coherence are a natural outcome of dynamical system views, as implemented in connectionist SRN models, where predictions are a result of self-organized language acquisition without an explicit representation of a grammar that ensures self-consistency.

In the meantime, a number of studies have provided evidence for effects of local syntactic coherence. The review of evidence and theories regarding these effects revealed a number of open questions that are worth further research:

Q 1 Are the results reported to be understood as resulting from local syntactic coherence – interference of more-word sequences – at all, or can they be explained by other means, such as, for example, lexical access?

Q 2 Are the results better explained as resulting from a modular system, implementing a preprocessing component that interacts with a later processing stage? Or do they rather provide evidence for an interactive system integrating different sources to guide action immediately?

Q 3 Do the effects mirror strategies that can be described as *rational* with regard to the task of sentence processing? Or do they merely reflect unwarranted, but unavoidable results of the inherent mechanisms of the processing system?

I will, unfortunately, not be able to provide conclusive answers to all aspects that are incorporated within these questions. However, in the remaining chapters I will evaluate three hypothesis that are, on the one hand, suited to gain insight into the phenomenon itself and, on the other hand, support its importance as a particularly interesting phenomenon if we are interested in the foundations of human language processing and the nature of cognitive mechanisms.

H 1 Local syntactically coherent sequences are processed and partially interpreted in parallel to the globally correct interpretation.

H 2 The influence is a result of frequent encounters with the locally coherent sequences in contexts where the local interpretation is the correct one.

H 3 The framework that can best account for the phenomenon is a dynamical system perspective as it is, for example, instantiated in simple recurrent network models of sentence processing.

To support the first hypothesis, I will revisit two visual-world experiments that provide particularly strong evidence that (a) local syntactic coherence does influence processing and that these influences cannot be accounted for by lexical retrieval processes (aiming at question **Q1**) and (b) that local syntactic coherence can trigger interpretational processes, which is, as I have argued, not easily compatible with modular approaches (**Q2**). The second hypothesis **H2** will be evaluated by complementing the visual world data with corpus data to assess the strength of local coherence within the different stimuli sentences. As I will show, the likelihood of encountering constitutive elements of the embedded sequence within contexts that support the local meaning modulates the effects. Finally, I will corroborate hypothesis **H3** by showing that both the influence of local syntactic coherence as well as the modulation of its strength arise in simple recurrent networks as a result of local predictions. Particular attention will be dedicated to evaluating whether the networks learn the underlying language in the first place. I will show that the pattern of effects cannot readily be explained as an artifact of a system that is not capable of grasping the crucial properties of a language, and would thus represent a rather poor cognitive model of human sentence processing. In contrast, effects of local syntactic coherence are an unavoidable outcome of the inherent mechanisms that enable the system to acquire and process language (**Q3**).

3 Interpretation of Local Syntactic Coherence: Two Visual World Experiments

In the following chapter I will report data from two visual world experiments that investigate the effects of local syntactic coherence. The basic ideas underlying these experiments have been developed over several years and in cooperation with a number of students and colleagues in the center for cognitive science at the University of Freiburg. The work presented here substantially benefitted from work that has been done in a series of project courses held by Lars Konieczny on connectionist models of language processing, which I attended as a student in 2002. As a student assistant in the winter terms of 2003/2004 and 2005/2006, and lecturer in winter 2008/2009, I was involved with the development of several experiments and modeling projects, including the two visual world experiments I will describe below. Stimulus materials of the experiments have been created mainly within these courses, and both experiments have been conducted by the students of the courses and a number of student assistants at our lab.

Experiment 1 was first presented as a conference poster at *AMLaP* (*Architectures and Mechanisms of Language Processing*) in Nijmegen in 2006 (Konieczny & Müller, 2006) and results have been published in the proceedings of the 31st Annual Conference of the Cognitive Science Society in 2009 (Konieczny, Müller, Hachmann, et al., 2009).

Experiment 2 was developed and conducted within the project *EloC* (*Empirical analysis of local coherence processing in sentence comprehension*) starting in 2007, funded by the German Research Foundation (*Deutsche Forschungsgesellschaft, DFG*; project KO 1932/3-1). Results of experiment 2 were first been presented at the 22nd Annual Meeting of the *CUNY* Conference on Human Sentence Processing, held at UC Davis (CA, USA) (Konieczny, Müller, Baumann, et al., 2009), and at a number of others conferences thereafter.

Both experiments are closely related and are a crucial source of evidence for the hypotheses put forward in this work. Data from experiment 2 will be used for the corpus-based analysis of frequency effects on local coherence processing (section 3.5) and also provide the basis for the connectionist model described in chapter 4. Therefore, I will describe the experiments in detail below and provide a extended statistical analysis.

As a first step, I will describe the visual world paradigm (section 3.1) and motivate why this experimental method is particularly suited to investigate effects of local syntactic coherence, and, in particular, to answer the question whether locally coherent sequences are being interpreted. I will outline the underlying idea of the two visual world experiments and spell out the hypotheses regarding local coherence processing (section 3.2). Experiment 1 (section 3.3) provides evidence that local syntactic coherence is not only processed on a abstract or syntactic level, but also influences interpretational processes. Experiment 2 (section 3.4) addresses some potential weaknesses regarding the design of experiment 1. Experiment 2 replicates the basic findings from experiment 1. In section 3.5, I will provide further support for the hypothesis that effects of local syntactic coherence can be explained as resulting from context-dependent distributional properties by showing that the corpus-derived tendency of the ambiguous elements to occur in the locally coherent meaning influences fixation patterns. Finally, in section 3.6, I will consider some potential issues regarding the method and materials used, and discuss the findings with regard to the different approaches described in section 2.3.

3.1 The Visual World Paradigm

In visual world experiments participants typically listen to auditory stimuli – sentences or instructions – and simultaneously watch (and sometimes manipulate) objects or scenes on a screen, while their eye movements are being tracked. Cooper (1974) was the first to use this method, and showed that eye movements are closely time-locked to the sentence unfolding over time. Extensive use of the paradigm, however, started only 20 years later, initiated by the publication of Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), who presented evidence that listeners use pragmatic cues in the visual input very early during online comprehension. Until then, the paradigm has been applied to a wide range of psycholinguistic topics, including sentence comprehension and lexical access, but also to production and acquisition. The visual world paradigm has been established as a very fruitful method to investigate online processes in language processing and the interaction of visual and linguistic information (see Huettig, Rommers, & Meyer, 2011, for a recent review). Eye movements as a window to mental processes have been used in reading experiments for a long period and have provided a rich source of behavioral data that are closely time-locked to stimuli presentation and thus enable, in contrast to offline methods, insights into online processes during language comprehension. Reading, however, is only one special case of language processing. With the visual world paradigm it is possible to use this rich source of information to investigate spoken language, and crucially, to access interpretational processes more directly than in reading experiments.

In addition, visual world experiments mimic a rather natural situation of language use. Very often utterances refer to entities in our visual environment, and an essential property of language use is to direct attention to particular aspects or objects in the environment. In a variety of studies it has been shown that, when listening to a sentence and being presented with visual input, the eyes quickly focus the mentioned or relevant entities in the course of the sentence. In addition, eye movements often reflect that we expect or anticipate particular items even before they were mentioned in the utterance (e.g. Altmann & Kamide, 1999), and has proved as a highly valuable tool to gain insights into interpretational processes during sentence processing and as a sensitive method to investigate which sources of information comprehenders can take into account while interpreting linguistic structures. Thus, the method provides a useful way of evaluating whether locally coherent sequences in the input string influence processing, and if such sequences trigger interpretational processes.

3.2 Using the Visual World Paradigm to Investigate Effects of Local Syntactic Coherence

Tabor et al. (2004) showed that comprehenders exhibit higher processing difficulty in sentences containing an LSC, compared to structurally similar sentences without LSCs. In addition, their second experiment provided evidence that also purely semantic differences lead to a similar effect, indicating that the effects are not solely based on syntactic features. However, differences in reading times can not readily test if or how deep an LSC has been *interpreted*. The visual world experiments described in the following sections attempt to clarify this question. The idea is the following: If comprehenders process local sequences independently of – or rather, in addition to – the global context, and if they interpret these local sequences at least partially, then a scene depicting the content of the local sequence should attract attention, mirrored in a higher likelihood of fixations on this scene. If so, this would provide further evidence for the influence of LSCs, and, crucially, favor interactive and parallel approaches that do not assign the effects reported in Tabor et al. (2004) to an encapsulated syntactic or lexical module that precedes semantic processing. Eye movements to the scene depicting the local content would have to be taken as strong evidence that interpretational processes are immediately triggered even if the global syntactic context does not support this interpretation, since it is not consistent with the syntactic structure of the input processed so far. Since any depicted content will attract some proportion of fixations, regardless of its match with the auditive stimuli, the effect of interest is the difference in proportion of fixations on a scene depicting the local context, dependent on the presence or absence of an LSC in the auditive stimulus.

Thus, we conducted two visual world experiments where participants listened to auditive stimuli either containing an LSC or not, while watching scenes depicting the content of the global sentence as well as the content of the LSC. In contrast to the majority of visual world experiments, we presented participants not only with one scene or an array of objects, but simultaneously with different scenes depicting local and global sentence meaning as well as a visually similar scene as unrelated control. Thus, it is possible to directly compare fixations on the different scenes between conditions, while keeping the visual display constant. In addition, since all three scenes contain the same set of referents, differences in fixation proportions are unlikely to be due to merely associative priming processes.[1]

Tabor et al. (2004) used reduced relative clauses in combination with morphologically ambiguous verb forms to construct sentences that contain the surface structure of a simple main clause and thus constitute an LSC. In German, however, the relative pronoun cannot be omitted, and so this exact type of LSC is not possible. Therefore, we have to rely on different sentence structures to evaluate the effects of LSC in German.

In contrast to English, subordinate clauses in German are verb-final and exhibit a *Subject-Object-Verb* structure, for example in complement clauses:

(30) Die Tatsache, dass die Astronautin den Außerirdischen entdeckt, entspricht der Wahrheit.
The fact, that the astronaut$_{NOM}$ the alien$_{ACC}$ discovers, corresponds with truth.
"The fact that the astronaut discovers the alien is true."

An adverb modifying the verb of the complement clause is usually located between the subject and the object of the subordinate clause:

[1] See Kukona et al. (2011) for an attempt to disentangle anticipatory processes based on the enfolding structure of the auditive stimuli and effects of merely associative semantic priming in visual world experiments.

(31) Die Tatsache, dass die Astronautin ungläubig den Außerirdischen entdeckt, entspricht der
The fact, that the astronaut$_{NOM}$ disbelievingly the alien$_{ACC}$ discovers, corresponds with
Wahrheit.
truth.
"The fact that the astronaut discovers the alien with disbelief is true."

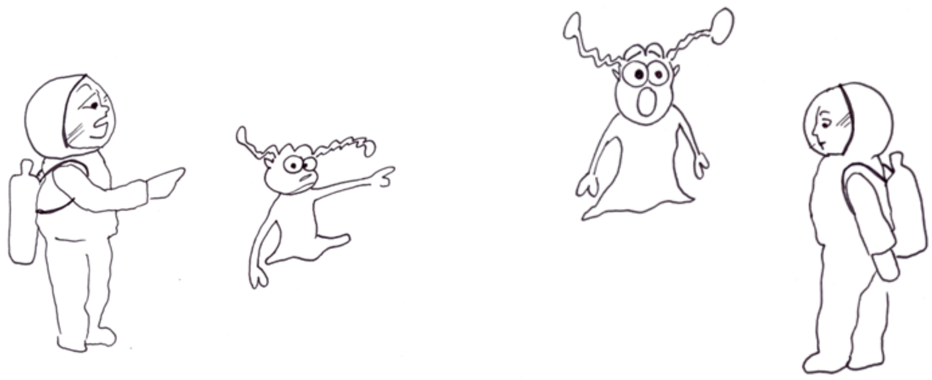
A number of tokens in German, for example «erstaunt» [amazes/amazed(ly)], are morphologically ambiguous as they can occur either as a finite verb in the third person singular («Sie erstaunt den Politiker» [She amazes the politician]), or as a participle («Sie ist erstaunt» [She is amazed.]). The participle can often be used as an adverbial element («Sie sieht erstaunt den Politiker» [She amazedly sees the politician.]). If such an ambiguous element is used as an adverbial element in a subordinate clause like (30), we can construct sentences that contain a sequence that is, regarding its surface, identical to a canonical simple main clause:

(32) Die Tatsache, dass [die Astronautin erstaunt den Außerirdischen] entdeckt, entspricht
The fact, that [the astronaut$_{NOM}$ amazes/amazedly the alien$_{ACC}$] discovers, corresponds
der Wahrheit.
with truth.
"The fact that the astronaut amazedly discovers the alien is true."

Importantly, taking into account the preceding sentence context, the finite verb reading of «erstaunt» [amazes] is not grammatically possible. Thus, we can construct sentences that contain a locally syntactic coherent sequence resembling a complete canonical main clause, similar to Tabor et al. (2004), which does not constitute a temporal ambiguity at any point in the sentence. A notable difference to the stimuli used in Tabor et al. (2004), however, is that in our stimuli the locally coherent sequence does not cross a clause boundary, but is embedded completely within the subordinate clause.

Crucially, the locally coherent sequence expresses a different proposition than the correctly parsed subordinate sentence, since the verbs of the local main clause structure and the subordinate clause differ. This makes it possible to present both propositions in the visual input: On the one hand, the astronaut that *amazedly discovers* the alien in figure 3.1(a), and on the other hand the astronaut that *amazes* the alien (figure 3.1(b)).

In the following section, I will describe two visual world experiments based on these considerations. While building on the same basic idea, in experiment 1 we used complement clauses as auditive stimuli, whereas object-extracted dative relative clauses were used in experiment 2 to address some potential weaknesses in experiment 1 that will be spelled out in the discussion of experiment 1 (section 3.3.4).



(a) global (b) local

Figure 3.1: Events depicting (a) the globally correct interpretation of the subordinate clause "the astronaut amazedly discovers the alien" and (b) the meaning of the locally coherent sequence "the astronaut amazes the alien"

3.3 Experiment 1

3.3.1 Design and Materials

Forty-eight German complement clauses were constructed, embedded in simple matrix clauses. Each complement clause contained a local syntactically coherent sequence constituted by an ambiguous participle («erstaunt» [amazes/amazedly] in example (33a)) in between the subject and the object of the complement clause. The word «erstaunt» is lexically ambiguous, as it can occur as an adverbial participle or a finite verb in the third person singular. Globally, «erstaunt» in (33a) must be analyzed as a participle-adverb [amazedly] of the clause-final verb «aufstöberte» [discovered]. Locally however, it constitutes the canonical *subject-verb-object* main clause «Die Astronautin erstaunt den Außerirdischen vom Mars» [the astronaut amazes the alien from Mars] as a full verb. The prepositional phrase «vom Mars» [from Mars] was added to prolong the locally coherent sequence, so that there's more time to build up and maintain a potential local subparse before the relative-clause verb completes the globally correct interpretation and thus "disambiguates" the sentence. As a control condition, the ambiguous participle was replaced by a synonymous unambiguous adverb («ungläubig» [disbelievingly / with disbelief] in (33b)), rendering the embedded substring ungrammatical and thus not constituting local syntactic coherence.

As a second control factor, the temporal adverb «gerade» [≈ just][2] was placed between the subject of the complement clause and the adverbial element (see examples (33c), (33d)). The added temporal adverb renders the local sequence incoherent, regardless of the ambiguity of the adverbial element, since the temporal adverb can only be placed after the verb in a main clause, but not before: «die Astronautin gerade erstaunt den Außerirdischen» is not a grammatical sentence of German.[3]

(33) Die Feststellung, dass …
The statement, that …
"The statement that …

a. *LSC: Ambiguous adverbial participle (+amb/-add.adv(LSC))*

… die Astronautin erstaunt den Außerirdischen vom Mars aufstöberte, …
… the astronaut$_{NOM}$ amazedly/amazes the alien$_{ACC}$ from mars discovered, …
… the astronaut amazedly discovered the alien from mars, …

b. *Control: unambiguous adverb (-amb/-add.adv)*

… die Astronautin ungläubig den Außerirdischen vom Mars aufstöberte, …
… the astronaut$_{NOM}$ disbelievingly the alien$_{ACC}$ from mars discovered, …
… the astronaut disbelievingly discovered the alien from mars, …

c. *Control: ambiguous adverbial participle, additional adverb (+amb/-add.adv)*

… die Astronautin gerade erstaunt den Außerirdischen vom Mars aufstöberte, …
… the astronaut$_{NOM}$ just amazedly/amazes the alien$_{ACC}$ from mars discovered, …
… the astronaut was just discovering amazedly the alien from mars, …

d. *Control: unambiguous adverb, additional adverb (+amb/-add.adv)*

… die Astronautin gerade ungläubig den Außerirdischen vom Mars aufstöberte, …
… the astronaut$_{NOM}$ just disbelievingly the alien$_{ACC}$ from mars discovered, …
… the astronaut was just discovering disbelievingly the alien from mars, …

… entsprach der Wahrheit.
… corresponded with truth.
… was true."

[2] In German, «gerade» in the particular context of the stimuli sentences expresses progressive, signifying that an agent is in the process of doing something: «Er liest gerade»: "He is reading." Thus, there's no good translation of «gerade» that preserves the syntactical structure of the sentence. As a compromise, I will use "just" in the translations.

[3] While the added temporal adverb eliminates the possibility of forming a complete locally coherent main clause within the complement clause, it introduces another locally coherent construction starting with a topicalized temporal adverb «Gerade erstaunt den Außerirdischen vom Mars (die Astronautin.)». Note, however, that this LSC lacks a subject-NP and is never actually completed because right after the object-NP «den Außerirdischen vom Mars» [the alien from mars] the finite verb «aufstöberte» [discovered] finishes the complement clause.

The second factor was included to address two potential issues: Firstly, differences in fixation proportions on the depicted events corresponding to the global and the local event might simply reflect the quality of the match between the visual properties of the depicted events and the two adverbs. If the addition of «gerade» [just] alters or even inhibits the effect, this possibility can be ruled out. Secondly, a similar logic holds for the evaluation of the unigram statistics model (Gibson, 2006, see 2.3.3.1), where any effect would have to be attributed to context independent lexical properties of the adverbial element, independent of the preceding context and thus regardless of the presence of a locally syntactic coherence. In the framework of this proposal, the presence or absence of «gerade» [just] should not elicit a systematic difference if the ambiguous element stays the same.

Tabor et al. (2004, p. 360) discusses this, or at least a rather similar explanation for their effects as *self-consistent parsing account 4* (SCPA 4). The fact that on encountering an ambiguous word both (or all) meanings are shortly activated can lead to higher processing-time compared to unambiguous words. Although the visual world paradigm does not measure processing difficulty directly, such an explanation would, as in Gibson (2006), predict effects of ambiguity regardless of the preceding context. Tabor et al. (2004) used the difference between reduced and unreduced relative clauses as a second control factor in their design, which, however, induces a considerable difference with regard to the constraining context preceding the ambiguous element, and may thus lead to less influence of ambiguity if the structural context strongly supports the globally correct interpretation. In contrast, the additional adverb «gerade» used here does not considerably change the constraints of the global sentence posed on the interpretation of the ambiguous element. Explanations based solely on lexical properties of the ambiguous element would, regarding the design used here, predict a main effect of ambiguity, but no influence of the additional adverb, since, in contrast to Tabor et al. (2004) the conditions are highly similar with regard to the syntactic structure and the constraining context preceding the relevant ambiguous element.

To sum up, materials were constructed according to a 2×2 design (table 3.1), comprising the factors adverb type *amb* (ambiguous +*amb* vs. unambiguous -*amb*) and coherency interruption by an additional adverb *add.adv* (without «gerade» -*add.adv* vs. with «gerade» +*add.adv*). Four lists were constructed such that each item appeared in one and only one condition on each list, according to a Latin square design. Each participant received one of the lists in randomized order.

Table 3.1: Design of visual world experiment 1. The first column displays the labels that will be used in figures and in the remainder of the text.

condition label	*amb*	*add.adv*	example			
			Die Feststellung, dass …			
+amb/-add.adv (LSC)	+	-	**die Astronautin**		**erstaunt**	**den Außerirdischen** …
-amb/-add.adv	-	-	die Astronautin		ungläubig	den Außerirdischen…
+amb/+add.adv	+	+	die Astronautin	gerade	erstaunt	den Außerirdischen…
+amb/+add.adv	-	+	die Astronautin	gerade	ungläubig	den Außerirdischen…
			… vom Mars aufstöberte, entsprach der Wahrheit.			

Auditive stimuli Auditive stimuli were recorded by a female speaker with a natural speech rate of about 43 words per minute (*SD* 2.62). In contrast to reading studies, in aural perception we have to deal with the problem that prosodic cues might weaken local coherence processing. To minimize the effects of prosody and differences in sound length due to differences in speech rate, locally coherent sequences were recorded separately as main clauses (34). Determiner and prepositional phrase were not taken from the main clause to avoid prosodic anomalies due to sentence-initial and sentence-final prosody. The remaining part (italic in (34)) was spliced into the spoken matrix sentences (35).

(34) main clause

Die ✂ *Astronautin erstaunt den Außerirdischen* ✂ vom Mars.
The ✂ *astronaut amazes the astronaut* ✂ from mars.

(35) resulting stimulus (33a)

Die Feststellung, dass die ✂ *Astronautin erstaunt den Außerirdischen* ✂ vom Mars
The statement, that the ✂ *astronaut amazes/amazedly the astronaut* ✂ from mars
aufstöberte, entsprach der Wahrheit.
discovered, corresponded to truth.

To minimize prosodic differences between conditions, control conditions were produced by splicing the critical words (unambiguous adverb and additional temporal adverb) into the target sentence (33a), as shown below.

(36) resulting stimulus (33b)

Die Feststellung, dass die *Astronautin* ✂ **ungläubig** ✂ *den Außerirdischen* vom Mars aufstöberte,
The statement, that the *astronaut* ✂ **disbelievingly** ✂ *the astronaut* from mars discovered,
entsprach der Wahrheit.
corresponded to truth.

(37) resulting stimulus (33c)

Die Feststellung, dass die *Astronautin* ✂ gerade ✂ *erstaunt* *den Außerirdischen* vom Mars
The statement, that the *astronaut* ✂ just ✂ *amazes/amazedly the astronaut* from mars
aufstöberte, entsprach der Wahrheit.
discovered, corresponded to truth.

(38) resulting stimulus (33d)

Die Feststellung, dass die *Astronautin* ✂ gerade ✂ **ungläubig** *den Außerirdischen* vom Mars
The statement, that the *astronaut* ✂ just ✂ **disbelievingly** *the astronaut* from mars
aufstöberte, entsprach der Wahrheit.
discovered, corresponded to truth.

Visual materials Each visual stimulus was composed of three different depicted events with the same pair of actors. One depicted event represented the globally correct content of the subordinate clause (here: «Die Astronautin entdeckt erstaunt den Außerirdischen» [the astronaut amazedly discovers the alien], figure 3.2a), the second picture depicted the content of the local coherence (here: «Die Astronautin erstaunt den Außerirdischen» [the astronaut amazes the alien], figure 3.2b), and the third picture showed the two actors in a unrelated interaction (such as both not taking notice of each other), neither matching the global nor the local content (figure 3.2c). In the remainder of the text, the three scenes will be referred to as *global*, *local* and *unrelated* scenes.

Since processing three, scenes each containing two interacting agents, is a rather complex visual task, depictions were held as simple as possible by using black and white hand drawings, avoiding unnecessary details as far as possible. A difficulty in depicting the meanings lies in the fact that the ambiguous elements that can be used within the sentence structures are mainly psych-verbs, that, more than signifying an action, signify the result of an action on the mental state of the patient/experiencer. Thus, depicting the locally coherent meaning is partly represented by the facial expression of the patient (e.g., the alien being amazed), and the globally correct, adverbial meaning in the facial expression of the agent (e.g., the astronaut being amazed and simultaneously doing something). The placement of the events on the screen was controlled, such that each event occurred equally often in each position.

Hypothesis A higher proportion of fixations is expected on the *local* scene depicting the meaning of the locally coherent sequence, when participants listen to a sentence that contains such a sequence (33a) than when listening to any of the control sentences ((33b), (33c) and (33d)). The effect is expected to show up shortly after the adverbial element has been heard and processed, which should have happened after about 200 ms after the word was presented (Allopenna et al., 1998). Regarding the duration of the effect, clear-cut expectations are diffi-

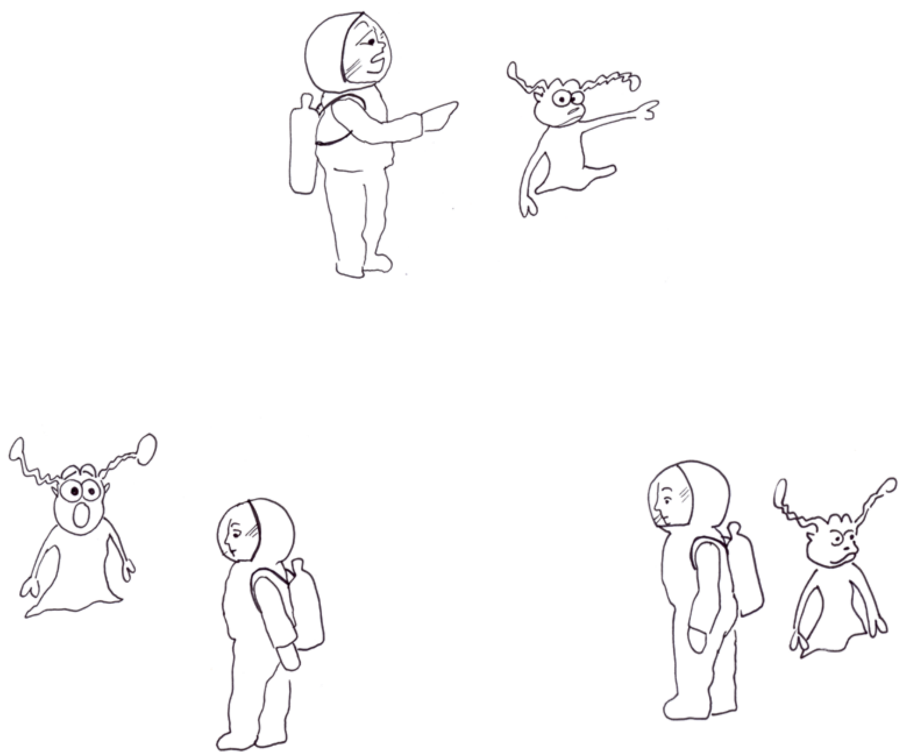


Figure 3.2: A visual stimulus with three depicted events: (a) *global* event (top), (b) *local* event (left), (c) *unrelated* event (right)

cult to formulate since little is known about how long a potential temporal local interpretation is maintained. However, it seems likely that the effect will vanish after the verb of the subordinate clause has been heard, since it provides the action that completes the globally correct proposition of the subordinate clause and thus “disambiguates” the local interpretation.

3.3.2 Procedure

Before the experiment started, participants read a description of the following procedure and were told that they had to judge the overall match of the pictures and the text used in the study in a questionnaire after the experiment. Each trial started with the presentation of the visual stimulus, followed by a short description of the depicted agents:

(39) Hier sieht man eine Astronautin und einen Außerirdischen in drei verschiedenen Szenen.
“Here you see an astronaut and an alien in three different scenes.”

After a pause of eight seconds the auditory stimulus was presented in its natural speech rate. The experiment did not include any explicit task, but was designed as a “look and listen” experiment: during the trials, participants had no explicit task but to listen to the sentences and look at the pictures.

48 participants, mainly students from the University of Freiburg and all native speakers of German, took part in the experiment that lasted about 45 minutes. Participants were paid € 7,50 or given course credits. Participants were presented 48 stimuli each (12 in each condition), as well as 24 filler items. Fixations on each of the three depicted events were recorded while the spoken sentences were played to the participants. Data was collected with an SR Research Eyelink II head-mounted eye tracker and an SR Research Eyelink 1000 desktop-mounted eye tracker in remote mode[4] sampling pupil position at 500 Hz.

3.3.3 Results

Interest areas were defined as three equally sized rectangles, identical for all trials, where each rectangle covers one of the three depicted scenes. Each fixation was mapped to the corresponding interest area. To visualize the time-course of gaze behavior, trial duration was split into timebins of 100 ms each. Proportion of fixation time on each of the three interest areas

[4]Since the interest areas cover a relatively large area, no stabilizing headrest was used. Although the remote mode leads to lower accuracy, it is sufficient for the purpose at hand, and provides a more comfortable setting for participants.

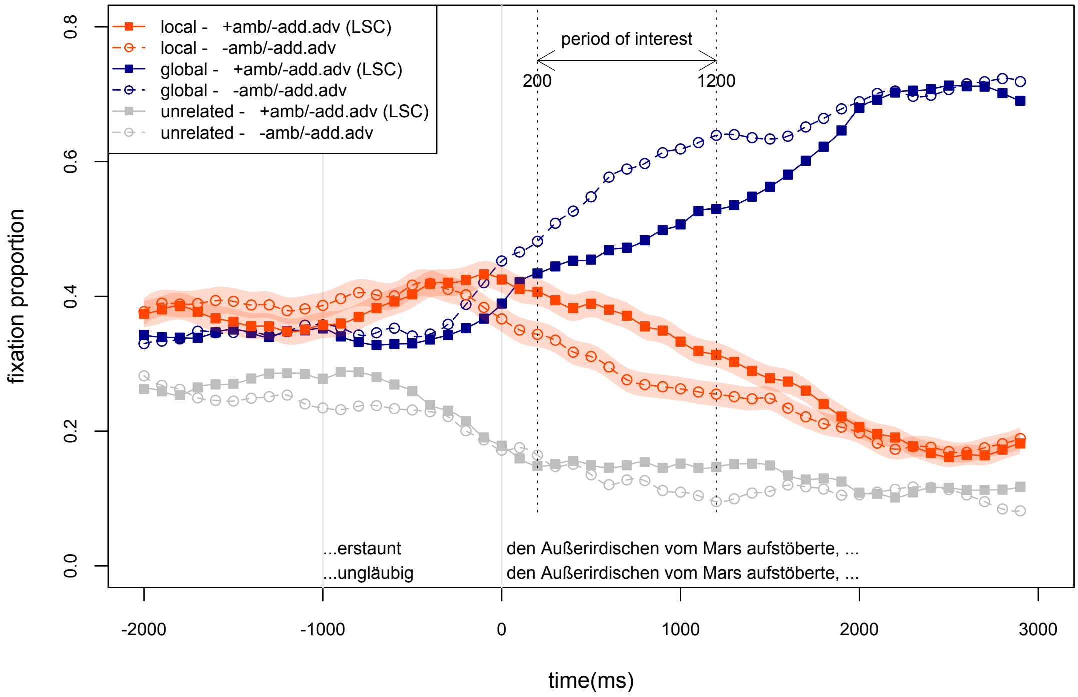


Figure 3.3: Fixation proportions in experiment 1. Conditions *+amb/-add.adv(LSC)* («erstaunt» [amazed/amazedly]) vs. *-amb/-add.adv* («ungläubig» [disbelievingly]); averaged over timebins of 100ms. Coloured bands signify standard errors.

(and other portions of the screen) was calculated for each timebin. To account for differences in stimulus duration, trials were synchronized at the offset of the ambiguous/unambiguous adverbial element, which is the point in time where the earliest influence of the experimental conditions is to be expected. Figure 3.3 to figure 3.5 show fixation proportions, plotted separately for each interest area and condition as a function of time. Dashed vertical lines signify the period of interest that was used for statistical evaluation later on.

The general pattern shows that participants tend to consistently fixate the *global* scene (blue lines) towards the end of the sentence, indicating that the depictions matched the meaning of the sentences quite well. Proportions of fixations on *global* and *local* scene (red lines) tend to separate at the offset of the adverbial element. The lower proportion of fixations on the *unrelated* scene (grey lines) over the complete time-range is likely to be due to the fact that in both *local* and *global* scenes the two agents are more active than in the *unrelated* scene and thus the visual salience of the *unrelated* scene is lower. The general tendency to direct their gaze to the globally matching scene as soon as the sentence allowed them to identify it, shows that participants are in general not distracted by the embedded locally coherent sequence in a way that they completely loose track of the correct interpretation. The relevant comparison with regard to our hypothesis, however, is the proportion of fixations on the *local* scene, dependent on the type of sentence participants hear.

Figure 3.3 compares proportions of fixations on the three interest areas when participants listened to a sentence with local syntactic coherence *+amb/-add.adv (LSC)* (33a) (solid lines) and when listening to a control sentence with an unambiguous adverbial element *-amb/-add.adv* (33b) (dotted lines). In spite of being on the right track, reflected in the higher fixation proportions on the globally correct scene (blue), participants pay more attention to the local scene when the sentence contains a locally coherent sequence, reflected in the distance between solid and dotted red lines. The separation starts roughly at the offset of the adverbial element and continues until about 1200 ms after the adverbial element has been heard. The unrelated scene does not attract much attention anymore at this time period, and does not show any considerable influence of the experimental manipulation. Rather, the embedded LSCs seem to temporally distract attention from the globally correct scene, visible in the difference between solid and dotted blue lines. The same pattern is visible when comparing conditions *+amb/-add.adv (LSC)* (33a) and *+amb/+add.adv* (33c), where the additional adverb «gerade» [just] inhibits local syntactic coherence (figure 3.4). However, since the duration of the parts of the stimuli preceding the synchronization point differs systematically between conditions due to the additional element, interpretation of this difference has to be done with caution.

Comparison of control conditions *+amb/+add.adv* (33c) and *-amb/+add.adv* (33d), both with additional «gerade» [just], does not reveal any visible difference in the period of interest (figure 3.5), indicating that the differences are not merely due to the presence or absence of an ambiguous element. However, there seems to be a rather late effect of ambiguity, starting at about 1200 ms after adverb offset and lasting until about 2700 ms, where fixation proportions on the local scene tend to be higher when the sentence contains an ambiguous element. We will give some tentative interpretations of this effect in the discussion section of experiment 1.

Statistical analysis To analyze the data, linear mixed effects modeling was applied (Pinheiro & Bates, 2000), using the statistical software *R* (R Core Team, 2013) and the *lme4* package (Bates, Maechler, & Bolker, 2013).

For statistical analysis, a period of interest was defined starting at 200 ms after the offset of the ambiguous element, allowing for processing of the word and saccade planning after

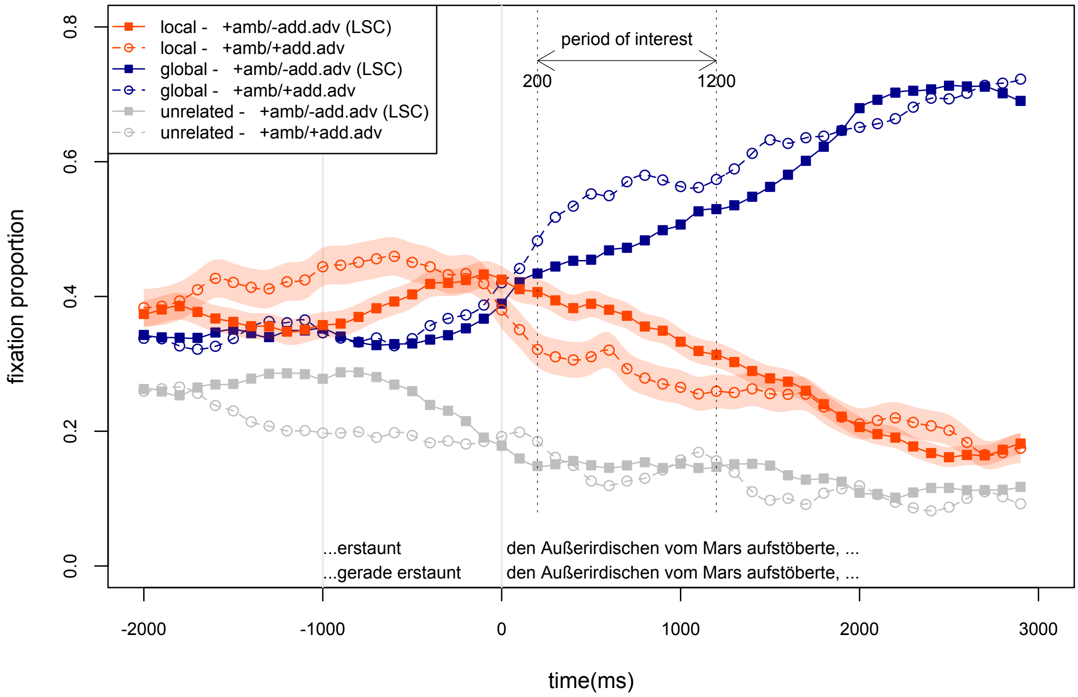


Figure 3.4: Fixation proportions in experiment 1. Conditions *+amb/-add.adv(LSC)* («erstaunt» [amazed/amazedly]) vs. *+amb/+add.adv* («gerade erstaunt» [just amazes/amazedly]); averaged over timebins of 100ms. Colored bands signify standard errors.

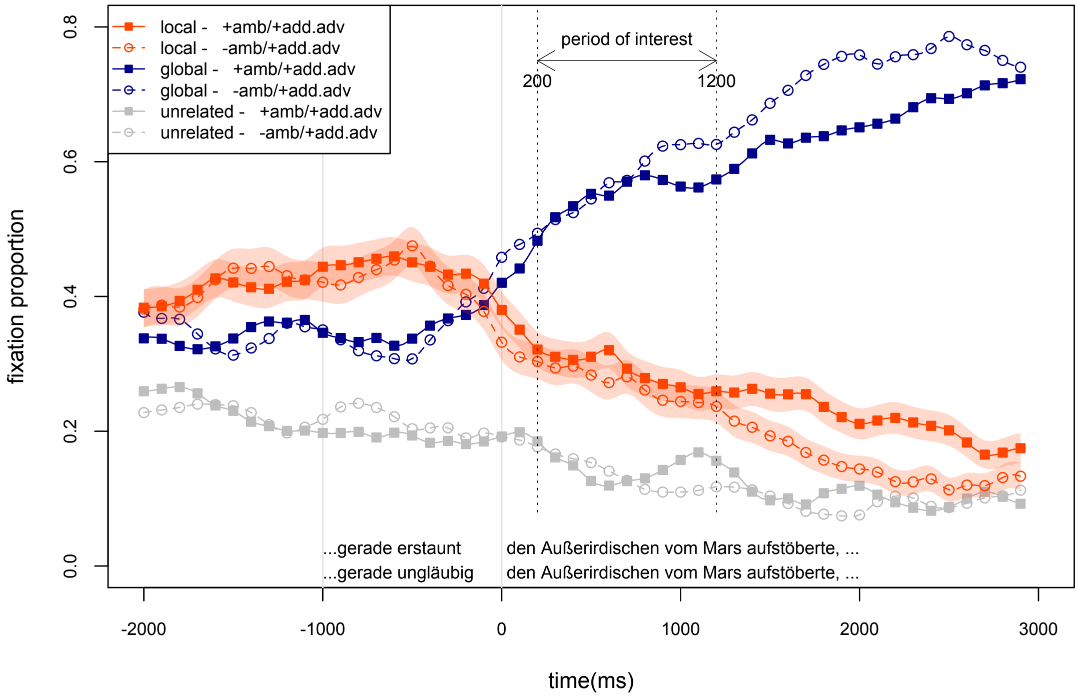


Figure 3.5: Fixation proportions in experiment 1. Conditions *+amb/+add.adv* («gerade erstaunt» [just amazes/amazedly]) vs. *-amb/+add.adv* («gerade ungläubig» [just disbelievingly]); averaged over timebins of 100ms. Bands signify standard errors.

hearing the word (Allopenna et al., 1998). The end of the period of interest was set to 1200 ms after offset of the ambiguous element. This rather long period allows for some delay in interpretational processes during the local coherent sequence, which ends with the onset of the relative-clause verb.

Calculating t-tests for the individual timebins that have been used to visualize the data in the preceding section is problematic for two reasons. Firstly, fixation proportions in the individual timebins can not be considered as independent events and thus violate the assumptions underlying the statistical method. Secondly, using short timebins and a relatively long period of interest has to cope with the problem of multiple comparisons. To overcome these problems, we calculated the proportion of fixations on each interest area for the whole period of interest, spanning 1000 ms. All fixations that occurred partly or fully within this time window were mapped to the interest areas, the resulting fixation time spent on each interest area during the period of interest was added up, and the proportion of fixation time on the local scene was calculated.

Fixation-proportions on the local scene were transformed to their logodds by calculating the logarithm of the ratio of fixation proportion on the local scene according to (3.1).

$$logit.poi.ia = log(\frac{p_{local}}{1 - p_{local}}) \tag{3.1}$$

The logit represents the probability to fixate the local scene compared to fixating any other location on the screen and provides an open range measure that allows us to use linear-regression modeling for proportional data.

The experimental factors *ambiguity* (*amb*) and *additional adverb* (*add.adv*) were coded as deviations from the mean of the binary (0,1) coding, such that positive values represent ambiguous adverbial elements (*+amb*, respectively the *absence* of an additional adverb *-add.adv*. This coding results in the easier interpretation of results, since the target condition containing a locally coherent sequence *+amb/-add.adv (LSC)* is associated with both positive values.

As Barr, Levy, Scheepers, and Tily (2013) showed in a series of tests with simulated data, linear-mixed effect models bear the danger of exhibiting high errors if the random-effect structure only includes random intercepts for participant and item. Thus, following Barr et al. (2013), the random-effect structure was kept maximal. Random slopes for both design factors as well as their interaction were included in participant and item terms.

The fitted model including both design factors and their interaction as predictors revealed no reliable interaction between the two factors ($Estimate = 0.4588$, $SE = 0.1649$, $t = 1.316$),

Table 3.2: Models of logit of fixation proportions on local scene; upper panel shows main effects and interaction, lower panel shows contrasts with reference to the target condition *+amb/-add.adv (LSC)*

Factor	Estimate	Std. Error	*t*-value
main effects and interactions			
Intercept	-1.6274	0.1477	-11.019
amb	0.4651	0.2468	1.884
add.adv	0.4263	0.1649	2.585
amb × add.adv	0.4588	0.3487	1.316
model formula (R syntax): logit.local.win ~amb * add.adv + (1 + amb * adv | participant) + (1 + amb * add.adv | item)			
contrasts: reference level: *+amb/-add.adv (LSC)*			
Intercept	-1.1753	0.2324	-5.057
-amb/-add.adv	-0.6185	0.2798	-2.211
+amb/+add.adv	-0.6565	0.2355	-2.788
-amb/+add.adv	-0.8157	0.2973	-2.744
model formula (R syntax): logit.local.win ~condition + (1 + amb * adv | participant) + (1 + amb * add.adv | item)			

although the effect points in the expected direction (see table 3.2, upper panel). However, since the main hypothesis is that participants spent more time on the local scene in the target condition (*+amb/-add.adv*) than in any of the three control conditions, a second model was fit to the data. Instead of the two design factors, condition was coded as a factor comprising 4 levels, where the target condition constitutes the reference level. Random effect structure was the same as in the preceding model. The second model revealed significant contrasts between target condition and all of the three control conditions (see figure 3.6 and table 3.2, lower panel).

In addition, models were fitted to subsets of the data each comprising only two of the four conditions, which were compared to base line models without the relevant design factor. Comparing conditions *+amb/-add.adv (LSC)* (33a) and *-amb/-add.adv* (33b) revealed a significant main-effect of Ambiguity ($Estimate = 0.6177$, $SE = 0.2863$, $t = 2.157$), and did significantly differ from the corresponding base line model ($p < .05$).[5] No effect of Ambiguity, however, was found when comparing *+amb/+add.adv* (33c) and *-amb/+add.adv*(33d), where local syntactic coherence is inhibited by an additional adverb ($Estimate = 0.1468$, $SE = 0.3175$, $t = 0.462$), and the model did not differ from the corresponding baseline model ($p = .6516$).

[5]Model comparison was conducted by likelihood ratio tests.

Similarly, comparing *+amb/-add-adv (LSC)* (33a) and *+amb/-add.adv* (33c) reveals a significant effect of *additional adverb* (absence) ($Estimate = 0.6594$, $SE = 0.2392$, $t = 2.757$), and a better model fit compared to the corresponding baseline model ($p < .01$). No effect of additional adverb showed up between conditions *-amb/-add.adv* (33b) and *-amb/+add.adv* (33d), where no ambiguous adverb is used ($Estimate = 0.1892$, $SE = 0.2325$, $t = 0.814$).

3.3.4 Discussion

When listening to sentences containing a local syntactic coherence, participants were more likely to fixate a scene depicting the meaning of an embedded local coherent sequence than when local coherence of the sequence was either inhibited by the use of an unambiguous adverb or by insertion of an additional adverb. The differences in fixation-proportion start right after the crucial ambiguous/unambiguous adverbial element and end about 1000 ms later. Although the statistical analysis did not reveal a significant interaction, contrasts between conditions provide evidence that the reason for the differences in fixation proportions are due to the local coherence, since no effects were found when comparing conditions without local syntactic coherence.

However, the additional adverb does not seem to totally eliminate the effect, but rather to produce a difference between «gerade erstaunt» (33c) und «gerade ungläubig» (33d), between roughly 1400 and 2700 ms after the offset of the ambiguous element (figure 3.5). On the one hand, this could be interpreted as evidence for purely lexical effects due to the ambiguity. However, since the early effect is clearly eliminated by the additional adverb, the later effect is most probably a second effect of local coherence. Firstly, the sequence «erstaunt den Außerirdischen», although not constituting a complete main clause sequence, could still lead participants to the interpretation of this sequence as a verb phrase with «erstaunt» [amazes] as finite main verb and «den Außerirdischen» [the alien] as object. This shorter and therefore possibly weaker local coherence could still lead to more fixations on the scene that depicts the astronaut amazing somebody (figure 3.2b). The delay could be ascribed to both the incompleteness of the sequence and the later onset of the local coherence (starting at «erstaunt» vs. starting at «die Astronautin»).

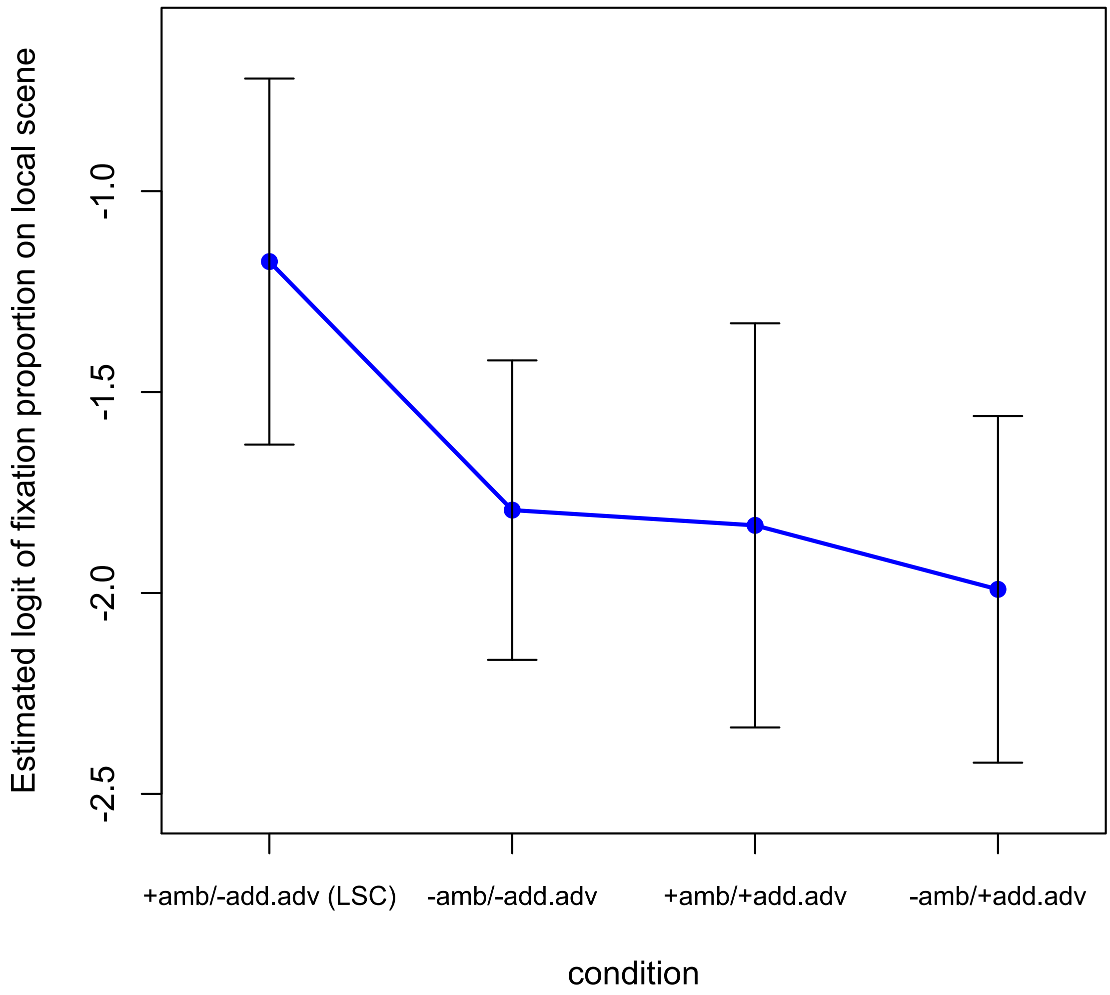


Figure 3.6: Estimated values of logit fixation proportion on local scene in experiment 1. Whiskers signify 95% confidence intervals

Secondly, the inserted adverb «gerade» can initiate a rather uncommon, but possible topicalized main clause such as

> Gerade erstaunt den Außerirdischen [die Astronautin]
> Just amazes the$_{ACC}$ alien$_{ACC}$ [the$_{NOM}$ astronaut$_{NOM}$]
> *"The astronaut is just amazing the alien"*

This possibility could further strengthen the interpretation of «erstaunt» as a finite verb, and thus trigger fixations towards figure 3.2b, where the alien is being amazed. However, since the effect is rather late after the synchronisation point, leading to growing noise due to differences in stimuli length after the adverbial element, these interpretations remain rather tentative. However, to reduce the likelihood of this potential influence, this issue will be addressed in experiment 2.

Another point can be seen as problematic in experiment 1. The ambiguous participles can, at least in principle, be interpreted as intransitive verbs in the context of the complement clause. The intransitive reading would result in a temporal globally consistent interpretation until the object NP of the embedded sentence has been heard:

> Die Tatsache, dass die Astronautin erstaunt, ...
> The fact that the astronaut amazes, ...

This interpretation would introduce a temporal ambiguity and could lead to a garden path effect as soon as the intransitive interpretation is disambiguated at the object NP «den Außerirdischen» [the alien]. The intransitive reading would, just as the locally coherent sequence does, reveal the astronaut as the subject of the finite-verb «erstaunt» [amazes]. If so, the effects would not tell us much about local coherence, but about garden-path processing. Two reasons, however, can be put forward against this explanation. Firstly, intransitive use of the ambiguous elements in the stimuli used is rare and in most cases sounds rather awkward, although it is, syntactically speaking, possible. Secondly, the intransitive reading is just as possible, or at least not more awkward, if the inserted adverb «gerade» is present:

> Die Tatsache, dass die Astronautin *gerade* erstaunt, ...
> the fact that the astronaut just amazes, ...

Thus, a potential garden-path due to temporal intransitive interpretation is not inhibited by the additional adverb and should therefore show up regardless of the presence or absence of the additional adverb. However, an effect of ambiguity was only present in sentences without additional adverb during the period of interest. The late difference between «gerade erstaunt» [just amazes/amazedly] (+*amb*/+*add.adv*, example (33c)) and and «gerade ungläubig» [just unbelievingly] (*-amb/+add.adv*, example (33d)), would in principle be compatible with a garden-path interpretation, but it seems to be hard to come up with an explanation that the inserted adverb should delay the garden-path interpretation for over a second. In addition, the intransitive meaning would obviously not include *the alien* as object, so that fixations on the local scene would solely be triggered by the subjecthood of the astronaut. However, since this possibility cannot not be ruled out completely, the potential temporal ambiguity will be avoided in experiment 2.

A third potential problem may be that the participle, in contrast to the unambiguous adverb, can be continued by an auxiliary, introducing either a passive or past perfect tense.

Die	Tatsache,	dass	die	Astronautin	erstaunt	→ *wurde*,	…
The	fact,	that	the	astronaut	amazed	→ *was*,	…

"The fact that the astronaut was been amazed … "

Die	Tatsache,	dass	die	Astronautin	erstaunt	→ *hat*,…	
The	fact,	that	the	Astronaut	amazed	→ has,	…

"The fact that the astronaut has amazed … "

While an expectation of a passive would match the global scene, where the astronaut is amazed, the past perfect matches the local scene, just as the intransitive reading mentioned above. However, if the difference in possible continuations of the partial sequence should lead to a systematic difference in gaze behavior, we would again expect an effect of ambiguity, regardless of the additional adverb.

However, since there is no clear main-effect of ambiguity, it is rather unlikely that the proposed alternative explanations can account fully for the observed pattern of fixation proportions. It may well be, however, that these issues are a considerable source of noise in the data, which may be one reason for the fact that, in spite of the clear numerical tendency, no significant interaction of the two design factors has reached significance.

The next section describes another visual world experiment using syntactic structures that do not constitute a potential garden path and make passive and past perfect interpretations rather unlikely. In addition, the potential local coherence induced by the additional adverb «gerade» [just] is eliminated by using a different additional adverb.

3.4 Experiment 2

The general design was similar to experiment 1, but to address the potential problems discussed in section 3.3.4, we used dative relative clauses instead of complement clauses. In dative relative clauses with singular masculine subjects, the case-marked relative pronoun «dem» [whom] signals that the first NP in the relative clause has to be followed by the dative-object *before* the finite (ditransitive) verb of the relative clause. This eliminates the potential confound with a garden path effect discussed above, since the ambiguous element can, just after the subject of the RC, not be interpreted as an intransitive verb.[6] The use of dative relative clauses, however, requires the introduction of a third agent as the subject of the main clause. To accomplish this without rendering the sentences and the visual stimuli much more complex, proper names were used as subjects of the main clause. Only masculine names were selected, because only then is the relative pronoun marked unambiguously as a dative. Since it seems awkward to present a sentence with visual stimuli where the subject of the sentence is not depicted, an additional face, representing the subject of the main clause, was presented in the middle of the screen, simultaneously with the three scenes. The short introduction that was played before the stimulus sentence was accordingly changed to:

(40) Hier sieht man Hugo, und in den drei Szenen eine Astronautin und einen Außerirdischen.

"Here you see Hugo, and in the three scenes you can see an astronaut and an alien."

This also makes clear that the proper name does not refer to one of the referents in the three depicted scenes.

3.4.1 Materials

48 sentences with embedded dative relative clauses were constructed. As in experiment 1, local syntactic coherence was constituted by a morphological ambiguous participle, which can only be interpreted as an adverbial participle in the global context of the sentence, but constitutes the local sequence of a canonical main clause if interpreted as a finite transitive verb.

[6]There are dative verbs in German («Hans, dem die Außerirdische gefällt/imponiert, ...») that could constitute a garden path in this position. However, the ambiguous elements used for the stimuli do not allow for such an interpretation.

As control, the ambiguous element was replaced by a synonymous unambiguous adverb, thus rendering the local interpretation ungrammatical. To control for context-independent, lexical effects, we included, as a second factor, an additional adverb before the ambiguous element, which also renders the local sequence ungrammatical. In contrast to experiment 1, the adverb «äußerst» [very/extremely] was inserted instead of «gerade» [just] for all items. «Äußerst» cannot be used to modify a finite verb, and thus can not constitute a possible beginning of a new sentence with a finite-verb reading of the ambiguous element, e.g., a topicalized main clause as in (41). The same sequence with «äußerst» instead of «gerade» results in an ungrammatical sequence:

(41) Gerade überrascht den Außerirdischen (die Astronautin)
Just surprises the$_{ACC}$ alien$_{ACC}$ (the$_{NOM}$ astronaut$_{NOM}$)
The astronaut just surprises the alien

(42) * Äußerst überrascht den Außerirdischen (die Astronautin)
* Extremely surprises the$_{ACC}$ alien$_{ACC}$ (the$_{NOM}$ astronaut$_{NOM}$)

As in experiment 1, this results in a 2 × 2 design, comprising the factors adverb-type (ambiguous vs. unambiguous) and coherency-interruption by an additional adverb (with «äußerst» vs. without «äußerst») (see table 3.3).

Table 3.3: Design of visual world experiment 2. The first column displays the labels that will be used in figures and in the remainder of the text.

condition label	*amb*	*add.adv*	example			
			Hugo, dem ...			
+amb/-add.adv (LSC)	+	-	**die Astronautin**		**überascht**	**den Außerirdischen** ...
-amb/-add.adv	-	-	die Astronautin		ungläubig	den Außerirdischen...
+amb/+add.adv	+	+	die Astronautin	äußerst	überrascht	den Außerirdischen...
+amb/+add.adv	-	+	die Astronautin	äußerst	ungläubig	den Außerirdischen...
			... vom Mars zeigte, hat so etwas noch nie gesehen..			

(43) Hugo, dem ...
Hugo, whom ...
'Hugo whom ...'

a. *LSC: Ambiguous adverbial participle*

...die Astronautin überrascht den Außerirdischen vom Mars zeigt, ...
...the astronaut$_{NOM}$ amazedly/amazes the alien$_{ACC}$ from mars shows, ...
'...the astronaut amazedly shows the alien from mars, ...'

b. *Control: unambiguous adverb*

…die Astronautin ungläubig den Außerirdischen vom Mars zeigt, …
…the astronaut$_{NOM}$ disbelievingly the alien$_{ACC}$ from mars shows, …
'…the astronaut disbelievingly shows the alien from mars, …'

c. *Control: ambiguous adverbial participle, additional adverb*

…die Astronautin äußerst überrascht den Außerirdischen vom Mars zeigt, …
…the astronaut$_{NOM}$ extremely amazedly/amazes the alien$_{ACC}$ from mars shows, …
'…the astronaut shows extremely amazedly the alien from mars, …'

d. *Control: unambiguous adverb, additional adverb*

…die Astronautin äusserst ungläubig den Außerirdischen vom Mars zeigt, …
…the astronaut$_{NOM}$ extremely disbelievingly the alien$_{ACC}$ from mars shows, …
'…the astronaut shows extremely disbelievingly the alien from mars, …'

…hat so etwas noch nie gesehen.
…has such thing yet never seen.
'…has never seen anything like that before.'

Four lists were constructed such that each item appeared in one and only one condition on each list, according to a Latin square design. Each participant received one list in randomized order. The auditive stimuli were generated and spliced in the same way as in experiment 1 (3.3.1), and were recorded by the same female speaker.

Visual materials Visual stimuli, as in experiment 1, consisted of three black and white drawings depicting scenes with two agents corresponding to the global and the local meaning and one scene with the same agents interacting in a way not related to the sentence. In addition, in the middle of the screen, the drawing of a face was depicted, signifying the subject («Hugo» in example (43)) of the matrix-clause (figure 3.7). The three scenes and the central face were displayed simultaneously, and the three scenes were rotated so that each scene appeared equally often in all three positions.

Clicking task In contrast to experiment 1, participants had to click on the scene depicting the matching content of the sentence they heard and they received feedback after each trial.

Hypothesis In contrast to experiment 1, the task to click on the matching scene makes it possible to evaluate effects of local coherence on accuracy as well as reaction times. Higher error rates, and, in particular, more clicks on the local scene are expected when participants listened to a sentence containing a locally coherent sequence (example (43a)) than when listening to any of the three control conditions. However, the clicking action has to be conducted long after the locally coherent interpretation has been ruled out by the remainder of the sentence.

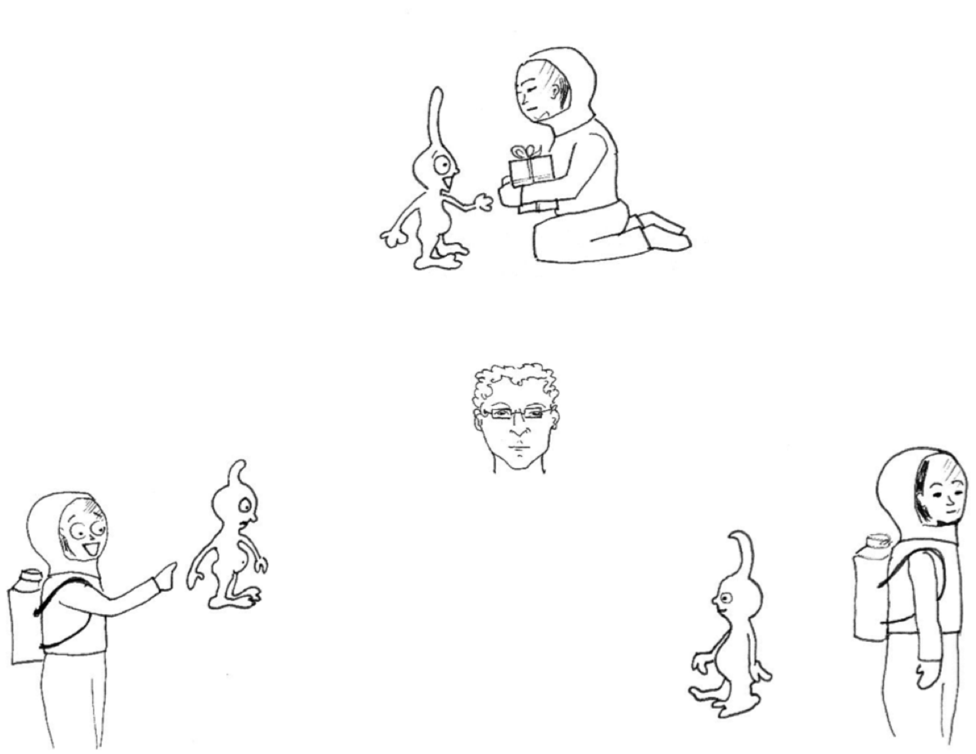


Figure 3.7: A visual stimulus with three depicted events and the subject of the matrix sentence: a. *global* event (top), b. *local* event (left), c. *unrelated* event (right), d. subject of matrix clause (center).

In addition, the results from experiment 1 showed that participants tended to consistently focus on the globally correct scene towards the end of stimulus presentation. Both factors may inhibit strong effects in the offline measures.

Regarding fixation proportion, hypotheses are similar to experiment 1. Processing of the locally coherent sequence should result in a higher proportion of fixations on the scene depicting the content of the LSC when participants listen to a sentence with LSC *+amb/-add.adv* (33a) than when listening to any of the control sentences, which should be reflected as an interaction of the two factors, *adverb ambiguity* and *additional adverb*. The effect is expected to should show up shortly after the adverbial element has been heard and processed. In contrast to experiment 1, the late effect of ambiguity in sentences with additional adverb that has been explained as possibly resulting from an additional, although shorter LSC induced by «gerade», should be eliminated in experiment 2, where «äußerst» has been used instead of «gerade».

3.4.2 Procedure

69 students of the University of Freiburg from different subjects took part in the experiment. All of them were native speakers of German and were between 19 and 32 years old (mean age: 24; female 45). Including additional tests, the session took about 45 minutes and participants were paid € 7,50 or given course credits.

Before the experiment started, participants read an instruction describing the procedure and were told that they had to click on the picture that matches the content of the sentence best after each trial. Each trial started with the presentation of a fixation point in the middle of the screen. Participants had to press the space-bar to start each trial. After pressing the button, the visual stimulus was presented, and after 5 seconds the short description of the depicted agents (example (40)) was played. After a pause of 4 seconds the auditory stimulus was presented. In contrast to experiment 1, participants had to respond to the sentence after each trial by clicking on the scene matching the sentence meaning. This task was included mainly to ensure that participants payed close attention to the stimuli, and to be able to evaluate accuracy in addition to gaze behavior. Admittedly, the additional task may change behavior significantly, but should rather force participants to focus on the correct scene more quickly. Thus, if an effect of local coherence can still be observed, it can be taken as even stronger evidence for the influence of "ungrammatical" interpretations than were provided in experiment 1.

One second after the auditive stimulus ended, the cursor appeared in the middle of the screen and a sound indicated that participants had to click on the scene matching the content of the sentence they heard. Participants started with 3 practice trials. Correct/incorrect feedback was provided after every trial and was displayed for 2 seconds. Participants were presented with 48 stimuli each (twelve in each condition), as well as 24 filler items. Fixations on each of the three depicted events and the central picture were recorded while the spoken sentences were played to the participants with an SR Research Eyelink 1000 desktop-mounted eye tracker sampling pupil position at 500 Hz (remote mode).

3.4.3 Results

Analysis of accuracy in the clicking task revealed a mean accuracy of 94.65%. In 2.7% of the trials participants clicked on the local scene, and on the unrelated scene in 2.6% of trials. However, for one item (item 2, see Appendix C.2) in only 57.4% of the trials participants clicked on the globally matching scene. This percentage is considerably lower compared to the rest of the items (all > 81.5%). In addition, in 23 of the 29 incorrect trials including the item participants clicked on the unrelated scene, indicating that the high error rate is most likely due to biases within the particular visual stimulus, and not, for example, the result of a particularly strong effect of local syntactic coherence.

Thus, this item (all 4 conditions) was excluded from further analysis. For the remaining items, accuracy was 95.45% (clicks on *local*: 2.62%, clicks on *unrelated*: 1.93%). All subjects showed accuracy above 85.11%, and all items above 79.41%.

3.4.3.1 Offline Measures

Reaction time Reaction times in the clicking task ranged from 373 ms to 23890 ms ($mean =$ 1856, $SD = 1836$). For further analysis of reaction times and accuracy, all trials with reaction times exceeding 6000 ms where excluded, which affected 123 trials, representing 3.84% of data points.

Reaction times where logarithmized to account for the skewed distribution of raw reaction times. To identify factors influencing reaction time regardless of the design factors, linear mixed effect models (with random intercept for item and participant) that did not include the experimental factors as predictors where fitted to the data. The following factors proved to reveal significant effects on log reaction time.

- position of the trial within the experimental session (*trial.pos*): The later a trial occurs in the session, the faster participants are to react. (*trial.pos*: $Estimate = -5.82e-03$; $SE = 3.79e-04$; $t = -15.35$)

- position of the global scene (*pos.global*): the screen position of the global (=correct) scene has a considerable influence on reaction time. If on top of the screen, participants react fastest. (*pos.global.top*: $Estimate = -1.16e-01$; $SE = 4.47e-02$; $t = -2.60$; reference level: *pos.global.left*)

- stimulus length (*stimulus.length*): the longer the auditive stimulus, the faster participants react.(*stimulus.length*: $Estimate = -5.80e-05$; $SE = 2.10e-05$; $t = -2.79$)

- correctness of answer: participants are faster if they click on the globally matching scene and thus give the correct answer, than when answering incorrectly by clicking on any of the non-matching scenes. (*answer.wrong*: $Estimate = 3.76e-01$; $SE = 4.22e-02$; $t = 8.92$)

Since the factor *add.adv* is constituted by an additional word in the sentence, stimulus length is confounded with this factor. Thus, stimulus length was not included as a predictor in further analysis.

To analyze whether local syntactic coherence influences reaction time, linear-mixed effect models where fitted to the data. The factors identified above, except stimulus length, where included as linear predictors. The factors *amb* and *add.adv* where coded by 0 and 1 such that the target condition with local syntactic coherence (*+amb/-add.adv (LSC)* (43a)) represents the reference level (+1,+1). Both factors where centered to the mean to account for missing values. Both factors and their interaction where included as predictors in the model. As recommended by Barr et al. (2013), both factors and their interaction where included as random slopes in both participant and item term.

Modeling revealed a significant main effect of *add.adv* such that participants where faster to respond if an additional adverb was included in the sentence (*add.adv*: $Estimate = 0.037$; $SE = 0.017$; $t = 2.20$). No effect of ambiguity and no interaction of the two factors proved reliable. All additional predictors identified with the simple base model also reached significance, and direction of effects was identical to the base model (see table 3.4).

Since answer accuracy revealed the largest effect, we fitted a similar second model, including only correctly answered trials. Again, only the main effect of *add.adv* was reliable, as well as the effects of the additional predictors (see table 3.5).

Table 3.4: Log reaction time, cutoff 6000ms

Factor	Estimate	Std. Error	*t*-value
Intercept	7.845	0.064	122.51
amb	-0.015	0.017	-0.87
add.adv	0.037	0.017	2.20
answer.correct	-0.380	0.0421	-9.02
pos.global.right	-0.007	0.042	-0.18
pos.global.top	-0.100	0.043	-2.33
trial.pos	-0.006	0.0004	-15.37
amb $\times$ add.adv	-0.004	0.031	-0.14

model formula (R syntax): log.rt ~amb * add.adv + pos.global + trial.pos + answer + (amb * adv | participant) + (amb * add.adv | item)

Table 3.5: Log reaction time, cutoff 6000ms, only correctly answered trials

Factor	Estimate	Std. Error	*t*-value
Intercept	7.462	0.051	146.60
amb	-0.0179	0.018	-1.02
add.adv	0.0376	0.0165	2.28
pos.global.right	-0.009	0.043	-0.20
pos.global.top	-0.103	0.044	-2.34
trial.pos	-0.006	0.0004	-14.99
amb $\times$ add.adv	0.0137	0.031	0.43

model formula (R syntax): log.rt ~amb * add.adv + pos.global + trial.pos + (amb * adv | participant) + (amb * add.adv | item)

Accuracy in the clicking task To analyze if local syntactic coherence influences accuracy in the clicking task, logistic regression models with crossed random effects where fitted to the data. As in the preceding section, influential linear predictors where identified by fitting base-line models that did not include the experimental factors. In contrast to reaction time, for accuracy only the position of the trial within the experiment revealed a significant effect (higher accuracy in later trials). ($Estimate = 0.0148$; $SE = 0.005$; $z = 2.902$; $p < .004$) and was thus included as a linear predictor.

As in the preceding section, trials with reaction times exceeding 6000ms where excluded. Coding of the experimental factors and random effect structure was identical. Similar to the results found for reaction times, only *add.adv* revealed a reliable main effect on accuracy (*add.adv:* $Estimate = -0.878$; $SE = 0.242$; $z = -3.627$, $p < .0005$), but not *amb*, and there was no interaction between the two factors. Trial position showed a significant effect such that participants became more accurate towards the end of the experimental session (see table 3.7).

Table 3.6: Answer accuracy, cutoff 6000ms

Factor	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	3.883	0.318	12.214	< 2e-16 ***
amb	-0.053	0.366	-0.145	.88
add.adv	-0.879	0.242	-3.627	< .0005 ***
trial.pos	0.019	0.006	3.335	< .001 ***
amb × add.adv	0.426	0.579	0.735	.462

model formula (R syntax): answer ~amb * add.adv + trial.pos +
(amb * add.adv | participant) + (amb * add.adv | item), family='binomial'

Clicks on locally matching scene Incorrect answers include both clicks on the local scene and the unrelated scene. Clicks on the unrelated control scene, however, are of little interest for our hypothesis since they cannot be clearly linked to potential LSC processing, but rather signify random errors. To evaluate if participants tend to click more often on the local scene when a locally syntactic coherent sequence is present in the auditive stimulus, trials with clicks on the unrelated scene (2.5% of the data) where excluded and logistic regression models with crossed random factors were fitted to the data. Click on the local scene (T/F) was used as the dependent variable, both design factors and their interaction were included as predictors as well as in the random terms for both item and participant. Coding of the experimental factors was as described in the preceding sections.

Again, only *add.adv* (and *trial.pos*) revealed reliable effects, such that participants were more likely to click on the local scene later in the session, and if an additional adverb was present.

Table 3.7: Clicks on local scene, cutoff rt < 6000ms, only trials where participants clicked either on the globally matching or the local scene.

Factor	Estimate	Std. Error	*z*-value	Pr(>|z|)
Intercept	-4.865	0.398	-12.211	< 2e-16 ***
amb	-0.169	0.491	-0.325	.74
add.adv	1.198	0.369	3.246	< .005 **
trial.pos	-0.0153	0.007	-2.088	< .05 *
amb × add.adv	0.209	0.715	0.292	.77

model formula (R syntax): click.local ~amb * add.adv + trial.pos + (amb * add.adv | participant) + (amb * add.adv | item), family='binomial'

Discussion of offline measures The results in the offline measures did not reveal any effects of local syntactic coherence. The main effect of *add.adv* in all measures, however, seems to provide evidence that the length of the stimulus plays a role for both reaction time and accuracy, probably by leaving more time to inspect the scenes and to plan the clicking action. In addition, screen position of the scene influenced reaction time rather strongly, which probably signifies that moving the mouse from the center of the screen to the top position is easier and faster than moving it to the left or right bottom. In general, participants show a slight training effect, both with regard to reaction time and answer accuracy.

3.4.3.2 Analysis of Gaze Behavior

For analysis of gaze behavior, no trials were excluded on account of reaction times for analysis of gaze behavior, since odd behavior in the clicking task is not necessarily reflected in the fixation patterns before attempting to answer. However, only correctly answered trials are included in the analyses.

To visualize the data, fixations were averaged over timebins of 100 ms, with sentences synchronized at the offset of the ambiguous/unambiguous adverbial element. Fixations were mapped to rectangular interest areas laid over the three scenes and the face in the center of the screen. Comparisons of fixation proportions between the relevant conditions are visualized in figure 3.8, figure 3.9, and figure 3.10. The general pattern resembles the results from experiment 1. Participants tend to consistently fixate the globally correct scene towards the end of

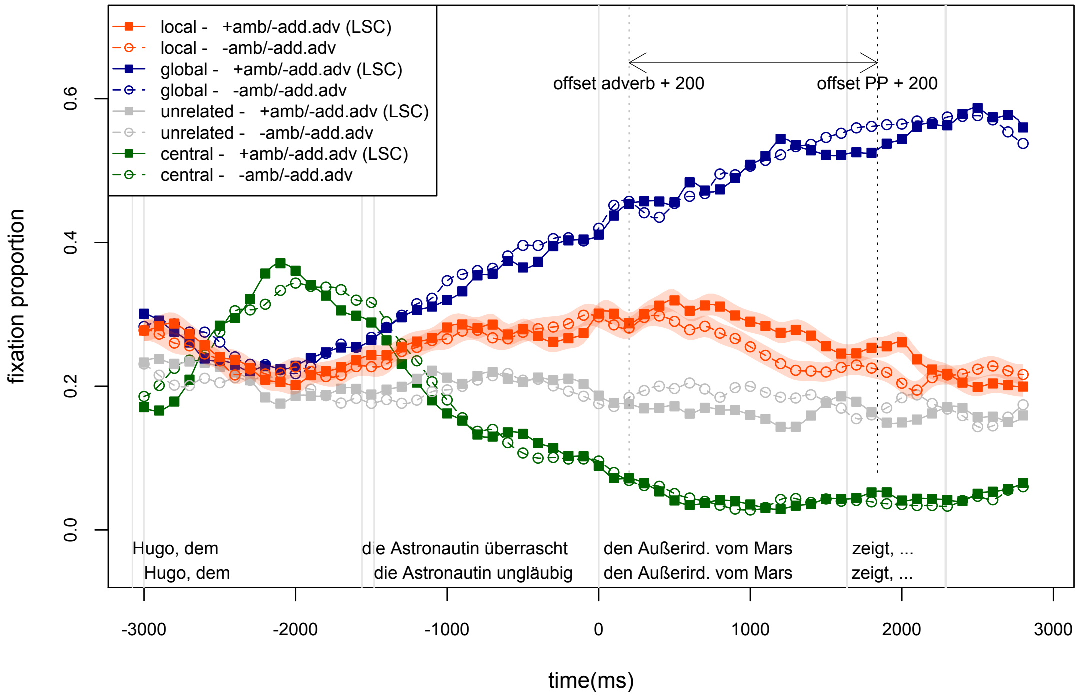


Figure 3.8: Fixation proportions in experiment 2. Conditions *+amb/-add.adv* («überrascht» [surprised/surprisedly]) vs. *-amb/-add.adv* («ungläubig» [disbelievingly]); averaged over timebins of 100ms. Bands signify standard errors.

the sentence, and spent less fixation time on the probably less interesting or salient unrelated scene than on both global and local scene. Fixations on the depicted face in the center of the scene representing the subject of the main clause drop quickly after the subject has been mentioned. Fixation proportions tend to be higher when participants listen to a sentence containing a locally syntactic coherent sequence than when listening to the control sentences with either an unambiguous adverb (figure 3.8) or an additional adverb inhibiting local syntactic coherence (figure 3.9). Ambiguity of the adverbial element alone does not seem to lead to any differences in the tendency to fixate the local scene (figure 3.10).

However, some notable differences in gaze behavior can be observed between experiment 1 and experiment 2. I will come back to these differences in the discussion (section 3.4.4).

Statistical analysis The period of interest was set as the time period spanning from the offset of the adverbial element until the offset prepositional phrase, marking the end of the

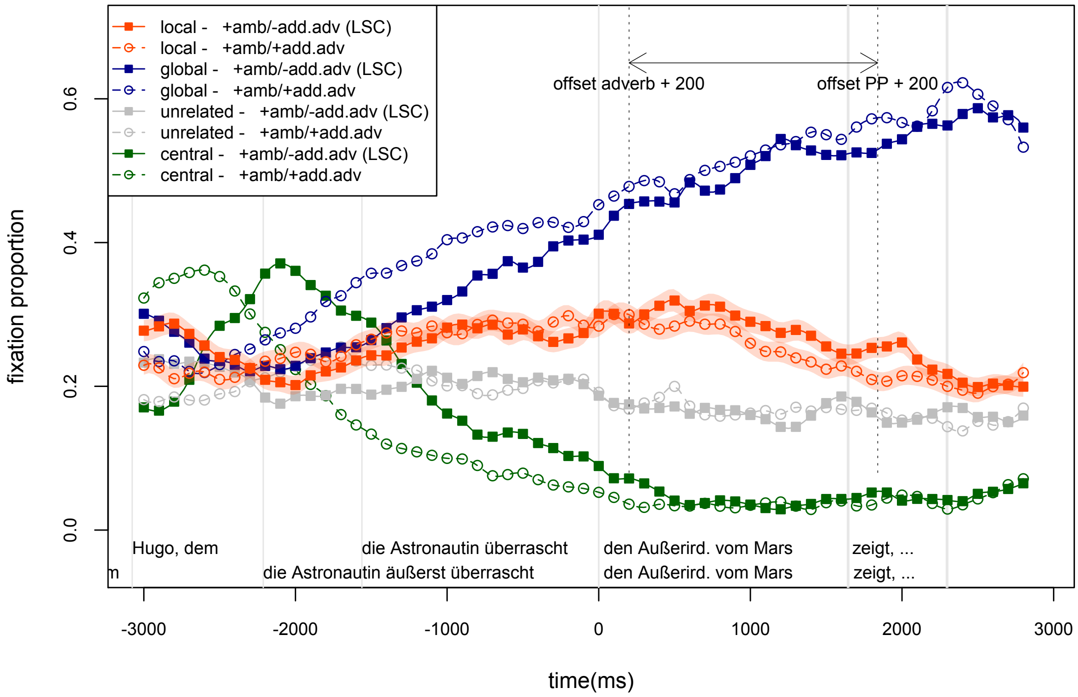


Figure 3.9: Fixation proportions in experiment 2. Conditions *+amb/-add.adv* («überrascht» [surprised/surprisedly]) vs. *+amb/+add.adv* («äußerst überrascht» [extremely surprised/-surprisedly]]); averaged over timebins of 100 ms. Bands signify standard errors.

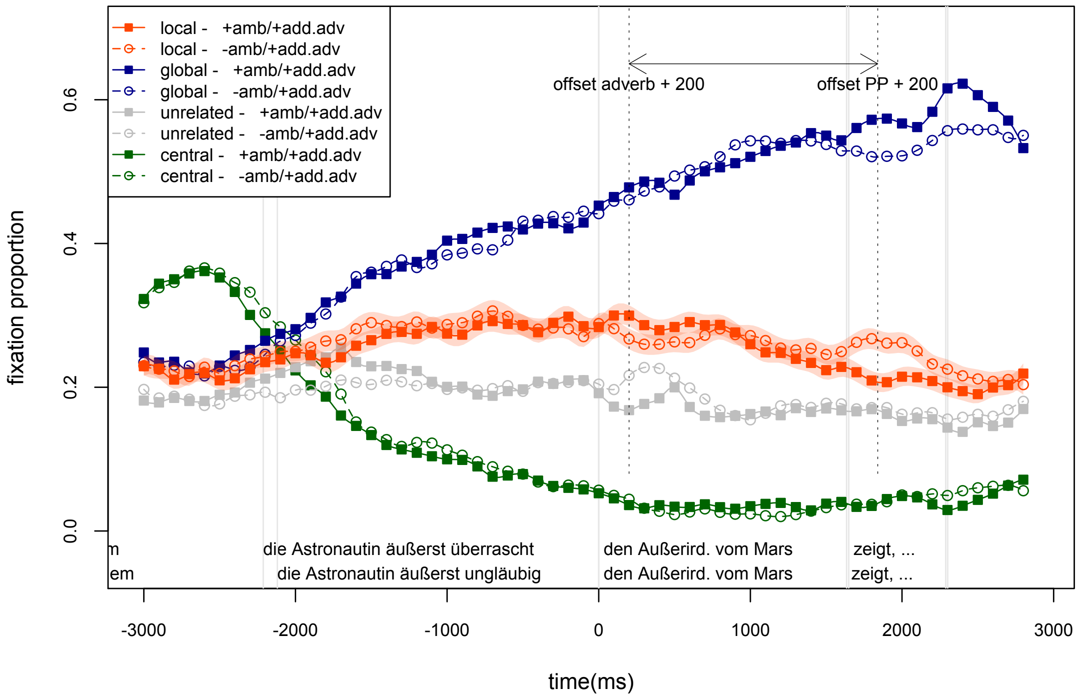


Figure 3.10: Fixation proportions in experiment 2. Conditions *+amb/+add.adv* («äußerst überrascht» [extremely surprised/surprisedly]]) vs. *-amb/+add.adv* («äußerst ungläubig» [extremely disbelievingly]]), averaged over timebins of 100 sms. Bands signify standard errors.

local syntactic coherence in the target condition (*+amb/-add.adv* (43a)). As in experiment 1, 200ms were added to both points in time to account for processing the word and saccade planning.

As in experiment 1, all fixations that occurred partly or fully within this time-window were mapped to the interest areas, the resulting fixation time spent on each interest area during the period of interest was added up, and the proportion of fixation-time on the local scene was calculated, and logits of fixation-proportions on the local scene constitute the dependent measure.

As in the analysis of offline measures (3.4.3.1), linear mixed effects models (with random intercepts for participant and item) were fitted to the data to identify factors that influence looking behavior independently of the experimental manipulation. The following factors proved to reliably influence the proportion of fixation time spent on the local scene during the period of interest (for statistical values see table 3.8):

- Visual salience of local scene (*logit.local.prestim*): To account for differences in visual salience of the different scenes, the logit of fixation proportions on the local scene in a time window of 10 seconds before the stimulus sound started was included as a linear predictor. The local scene tends to be fixated more if participants have spent more time on that scene before the stimulus sound started.

- Proportion of fixations on local scene in previous time window (*logit.local.prewin*: Another significant predictor proved to be the logit of fixation proportion on the local scene in the timebin before the period of interest. A time window was chosen spanning from 200 ms from onset of the LSC (= onset of RC subject) until the start of the period of interest. More time is spent on the local scene if more time was spent on the scene in the preceding time window.

- Scene position (*pos.global*): In the analysis of reaction times, a considerable effect of screen position was observed. It turned out that in general, a scene is more likely to be fixated if positioned on top of the screen than in another position. The strongest predictor for fixation proportions on the local scene turned out to be the position of the globally matching scene. Position of global scene was therefore included as a three-level factor (top, left, right).

- position of the trial within the experiment (*trial-pos*): As in the offline measures, trial position has a small, but reliable effect. Interestingly, it turned out that trial position reliably interacts with the *age* of the participants. While fixations on the local scene decreases over trial for younger participants, this is less so for older participants. Trial position revealed an independent main effect, but not age. Thus, trial position and the interaction of age and trial position were included in the model.

Table 3.8: Base model of logit of fixation proportions on local scene; only correctly answered trials. *p*-values are derived by MCMC sampling

Factor	Estimate	Std. Error	*t*-value	p_{MCMC}
Intercept	-1.645	0.211	-7.798	.0001
trial.pos	-0.052	0.013	-3.882	.0002
logit.local.prestim	0.074	0.048	1.519	.1216
logit.local.prewin	0.036	0.016	2.210	.0288
pos.global.right	-0.349	0.222	-1.572	.1186
pos.global.top	-0.607	0.226	-2.683	.0082
age × trial.pos	0.002	0.0005	3.219	.0010

model formula (R syntax): logit.local.win ~age : trial.pos + trial.pos + logit.local.prestim + logit.local.prewin + pos.global + (1 | item) + (1 | participant)

Given this baseline model, condition was included as a further predictor. Since the hypothesis is that fixation proportions on the local scene in the target condition (*+amb/-add.adv*) should be higher than in any of the three control conditions, contrast coding was used and the target condition was set as the reference level. Thus, negative estimates and *t*-values signify lower logits than in the target condition containing an LSC. Both design factors and their interaction were included in both random participants and random items terms. The model revealed lower fixation proportions in any of the three control conditions (figure 3.11), significant at the 95% level, given the directedness of the hypothesis.

Currently, no generally accepted method is available to derive *p*-values for models with random correlations. Thus, *p*-values were derived for a simpler model including only random intercepts for both participant and item by MCMC sampling [7]. In this anti-conservative model, all contrasts exhibited *p*-values < .05. Note, however, that *t*-values for the simple model are slightly higher than for the complex model, and that the random intercept only model tends to be anticonservative.

Relation to other portions of the sentence Analyzing only one particular period of interest, however, does not provide any information about what happens before and after this

[7] The function *pvals.fnc* was used, which is provided in the *languageR* package (Baayen, 2009)

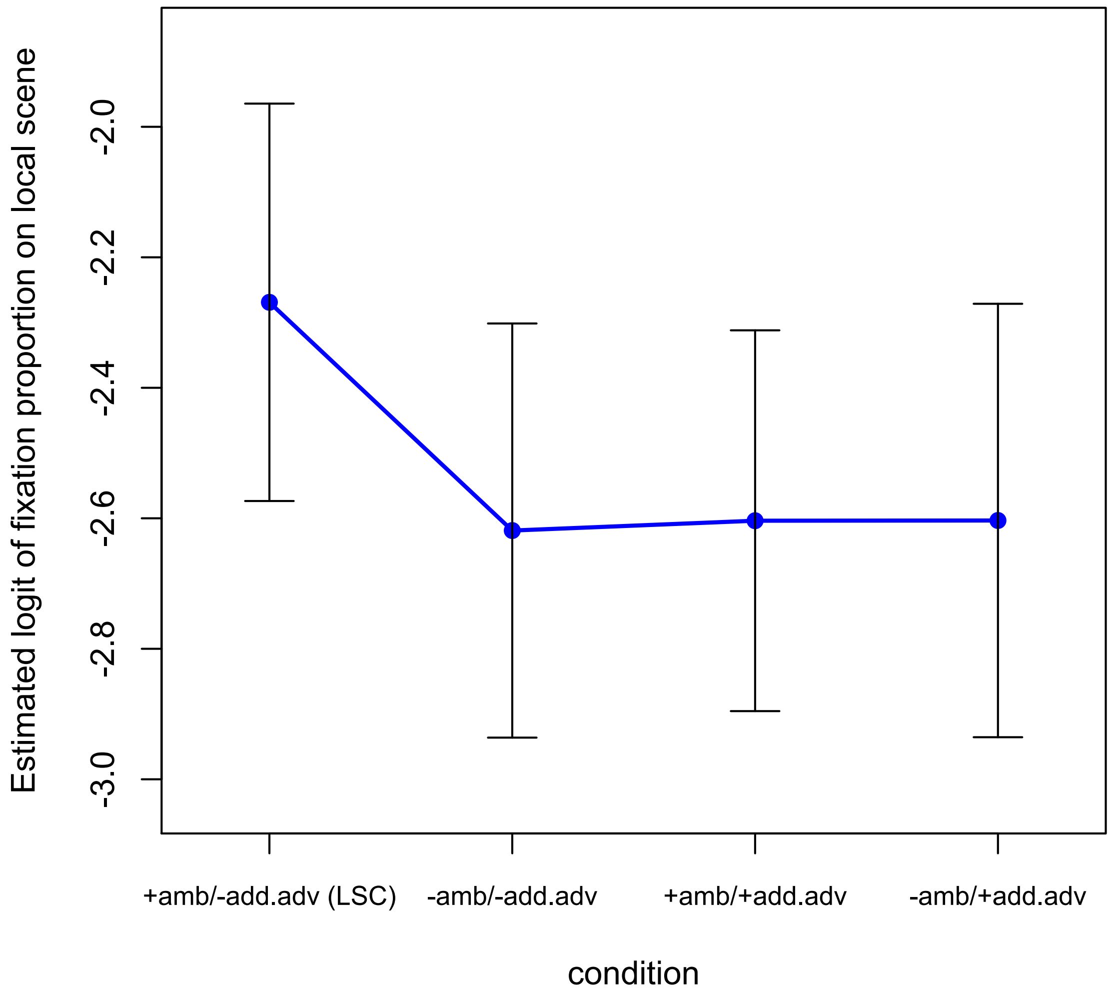


Figure 3.11: Estimated logits of fixation proportion on the local scene in experiment 2. Whiskers signify 95% confidence intervals.

Table 3.9: Models of logit of fixation proportions on local scene; only correctly answered trials. p-values for the simple model are derived by MCMC sampling

Factor	Estimate	Std. Error	t-value	p_{MCMC}
complex model, maximal random effects structure				
Intercept	-1.361	0.239	-5.689	
-amb/-add.adv	-0.350	0.183	-1.908	
+amb/+add.adv	-0.335	0.197	-1.748	
-amb/+add.adv	-0.335	0.199	-1.686	
trial.pos	-0.053	0.013	-3.973	
logit.local.prestim	0.074	0.048	1.533	
logit.local.prewin	0.035	0.016	2.168	
pos.global.right	-0.340	0.220	-1.815	
pos.global.top	-0.636	0.224	-2.836	
age × trial.pos	0.002	0.0005	3.311	
simple model, random intercepts only				
Intercept	-1.3871	0.235	-5.894	.0001
-amb/-add.adv	-0.360	0.172	-2.092	.0412
+amb/+add.adv	-0.340	0.172	-1.992	.0460
-amb/+add.adv	-0.343	0.172	-1.998	.0458
trial.pos	-0.052	0.013	-3.841	.0002
logit.local.prestim	0.072	0.048	1.480	.1324
logit.local.prewin	0.037	0.016	2.264	.0224
pos.global.right	-0.352	0.222	-1.584	.1206
pos.global.top	-0.610	0.226	-2.693	.0086
age × trial.pos	0.002	0.0005	3.189	.0014

period of interest. Contrasts between conditions during the period of interest can only be taken as stemming from the presence or absence of an LSC if this contrast exhibits a considerably different pattern than in other time windows, in particular during parts of the sentence where conditions do not differ from each other in relevant aspects. Therefore, I also analyzed other time windows. Since the ambiguous element is the point in time at which local coherence starts, no difference between conditions should be found before this point. Since the verb of the relative clause does not fit into the locally coherent parse, it is also likely that potential effects of a local interpretation should vanish when the RC verb has been heard.

Table 3.10: Periods of interest in experiment 2

	period of interest			
condition	preloco	beginloco	loco	afterloco
+amb/-add.adv (LSC)	Hugo, dem	die Astron. überrascht	den Außerird. vom Mars	zeigt, ...
-amb/-add.adv	Hugo, dem	die Astron. ungläubig	den Außerird. vom Mars	zeigt, ...
+amb/+add.adv	Hugo, dem	die Astron. äußerst überrascht	den Außerird. vom Mars	zeigt, ...
-amb/-add.adv	Hugo, dem	die Astron. äußerst ungläubig	den Außerird. vom Mars	zeigt, ...

I fitted linear mixed models similar to the one described above, differing only in the dependent variable (logit of fixation proportion on the local scene in the period of interest) and in the particular timebin representing the preceding period. Figure 3.12 shows the estimated logits [8] of fixating the local scene in the 4 periods of interest. Colored bars represent the three control conditions, plotted as contrasts to the condition containing a local syntactic coherence, which has been set as the reference level in the models. While no significant difference shows up in *preloco*, in *beginloco* the local scene is more likely to be fixated if an additional adverb is present. This main effect is likely to be due to the different length of this portion of the sentence if an additional adverb is present. Importantly, in *loco*, more fixation time is spent on the local scene when a local syntactic coherence is present, than in any of the three control conditions. In *afterloco*, logits are significantly higher in condition 4. Admittedly, I do not have any reasonable explanation for this particular effect.

Experiment 2 replicates the general patterns from experiment 1, showing that participants' eye movements are influenced by locally coherent sequences. The scene depicting the meaning of the locally coherent sequence, which is not compatible with the global context of the sentence, nevertheless attracts participants' eye movements.

[8] Estimates are derived with the *Effect* function, available in the *effects* R-package (Fox, 2003).

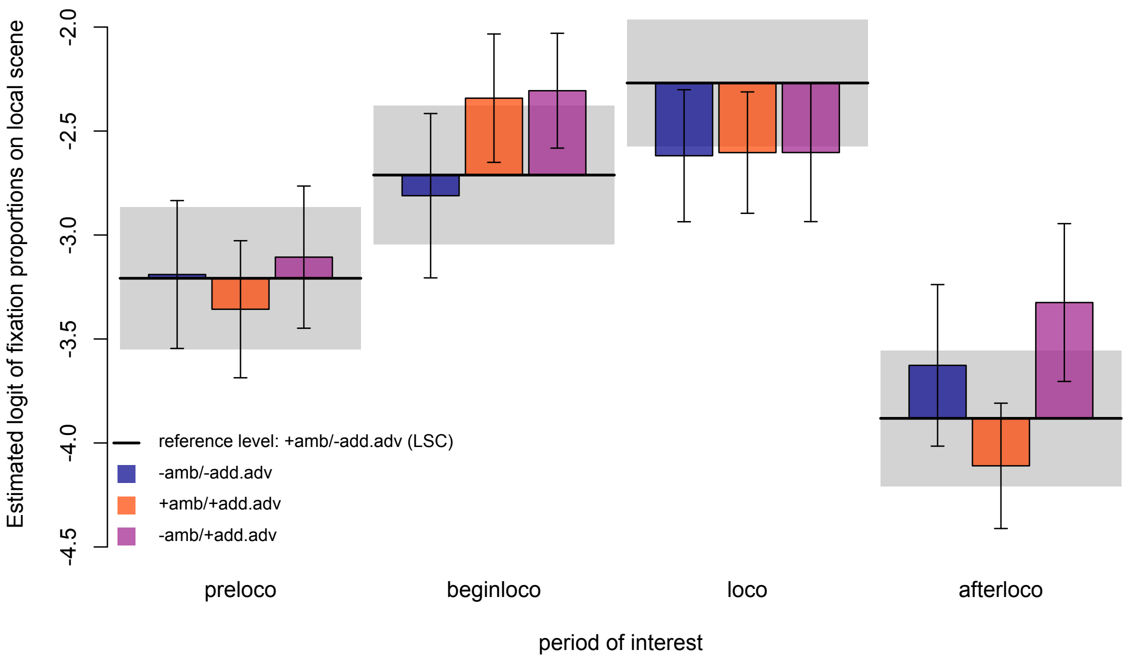


Figure 3.12: Estimated logits, plotted as contrasts of the 3 control conditions to the condition containing a local syntactic coherence, in different regions of the auditive stimulus. Whiskers signify 95% confidence intervals, the shaded rectangle represents 95% confidence interval of the intercept.

Offline measures as reaction times and accuracy in the clicking task did not reveal an effect of local coherence, but only an effect of the presence or absence of an additional adverb, which is most likely a result of the difference in stimulus length due to the additional word. In contrast to experiment 1, we used dative relative clauses to avoid a potential confound with garden-path effects, and used a different additional adverb that does not constitute the potential beginning of a shorter locally coherent sequence.

3.4.4 Differences Between Experiments 1 and 2

Although the effects did show up similarly in both experiments, there are some notable differences. The late effect (verb-coherence) observed in experiment 1 disappears in experiment 2, providing evidence that the «gerade» did indeed initiate the processing of a local coherence in experiment 1, while the «äußerst» in experiment 2, since it cannot begin a sentence, did not elicit this effect.

In experiment 1, the increase of fixation proportions on the global scene started roughly at the offset of the adverbial element (see, for example, figure 3.3), while in experiment 2 the decision to focus on the globally correct scene seems to start much earlier. As can be seen in figure 3.8, the blue and red lines tend to separate right after the relative pronoun has been heard. This difference in gaze behavior can most likely be attributed to the different sentence types used in the experiments, and, as a consequence, the type of verbs that are required in the particular sentence types. While the complement clauses require a transitive verb, the dative relative clauses require a ditransitive verb. Thus, the dative relative pronoun provides a more constraining context than the complementizer with regard to the particular verb. The use of dative relative clauses, starting with a dative-marked relative pronoun, may strongly trigger the expectation of ditransitive actions like «zeigen» [show], «geben» [give], and, due to the design of the experiment, only the globally correct scene depicts such an action, while the local scene depicts a transitive action («überraschen» [surprise], «erstaunen»[amaze]). Since participants have considerable time to inspect the scenes and encode the particular actions that are depicted in the scenes, it is not unlikely that a dative relative pronoun provides valuable cues about which scene will probably be referred to. More generally, the dative relative clause provides a much more constraining context on the class of verbs, and, consequently, actions that can possibly be described within the relative clause. As has been mentioned in section 3.3.4, until the adverbial element has been processed, a variety of structures is possible in the complement clauses, including passive, past perfect, and intransitive verbs, which are compatible with different scenes. In the case of a dative relative clause, the only liable verb class is ditransitive, the recipient is obligatory before the verb occurs, and so on. Thus, it may be much easier to figure out the matching scene relatively early in experiment 2.

Another reason, of course, may lie in the explicit task in experiment 2. While in experiment 1 participants only were advised to look and listen, the task to click on the correct scene required in experiment 2 may have motivated participants more strongly to try to identify the matching scene quickly.

Despite this deviation in gaze behavior, the differences between conditions closely match the results from experiment 1. Comparing sentences with local syntactic coherence and sentences with an unambiguous adverb again shows a difference such that participants are more likely to fixate the local scene when listening to a sentence containing local syntactic coherence (figure 3.8). The same holds true when comparing sentences with and without an additional adverb (figure 3.9), showing that the additional adverb inhibits looks to the local scene. No differences, however, are visible when comparing conditions without local coherence (figure 3.10). In contrast to experiment 1, the late effect of ambiguity in sentences with additional adverb (figure 3.5) does not show up in experiment 2 (figure 3.10).

3.5 Frequency Effects in Local Coherence Processing

The visual world experiments reported in the preceding sections provide further evidence that local syntactic coherent sequences indeed influence processing, as has been shown in self-paced reading experiments by Tabor et al. (2004) and others. The fact that in the visual world experiments the differences in fixation proportions can be best explained if the local sequence is at least partially interpreted and thus attracts fixations to a scene depicting a proposition that differs from the role-assignment compatible with the global sentence. So far, local syntactic coherence has been treated as a categorical factor, which is either present or absent, operationalized by using ambiguous or unambiguous words, or by slightly changing the embedded sequence in a way that it is not longer coherent. In terms of the connectionist approach adopted here, as well as in the majority of other proposals to explain the effects, however, local syntactic coherence is conceived as resulting from particular distributional properties in the linguistic environment and is thus considered as a gradient property. As Bicknell et al. (2009, p. 6) put it, a locally coherent sequence is "a string of words *w* that out of context would suggest one very likely parse, and that parse is impossible (or at least highly unlikely) in context."

Thus, sequences in a sentence can be more or less locally coherent, and, consequently, stronger local coherence should lead to stronger interference, competition or require more effort to reallocate resources. Within the dynamical system approach adopted here, the strength of the attractor built up by the local sequence will depend on the likelihood of the local parse in other contexts, and thus on how often the system has experienced the sequence (or relevant elements of the sequence) in contexts where the local parse has been appropriate. Conceiving the system as moving continuously through a constantly changing attractor landscape shaped by linguistic experience and current input, even slight differences in the distributional patterns will influence the trajectory of the system, leading to a more or less large detour on its path.

With regard to the visual-world experiments reported above, the locally coherent sequence is constituted by morphologically ambiguous elements that allow for a finite verb reading that is inconsistent with the preceding sentence context. Ambiguity in general is hardly to be considered as a categorical factor; ambiguous elements can be more or less ambiguous, and, crucially, the elements will typically be strongly biased towards one of the different alternatives. If we think about the usage of «verschüchtert» [intimidat(ed)], we will most likely make up a sentence where somebody is «verschüchtert» or acts «verschüchtert». In contrast, thinking of «überrascht» [surprise(d)], this may first trigger a scene where somebody, or something, surprises me. This intuition will most likely be closely related with our experience with the particular forms in the language. Evidence that comprehenders are sensitive to the distributional properties of ambiguous elements has been shown in a range of studies, in particular with regard to ambiguity resolution, and has been a major ingredient of the constraint-based interactive framework described in section 2.1.1.4.

With regard to the phenomenon investigated here, local coherence should be stronger and therefore more distractive if we consider the ambiguous element more as a verb rather than as an adjective or adverb. On the other hand, if we encounter a word mostly or only in contexts where it does not function as a finite verb, the language processing system may not even notice the – in principle possible – alternative usage and will thus be less distracted.

3.5.1 Estimating the Strength of Local Syntactic Coherence

Local syntactic coherence in the auditive stimuli in both visual-world experiments is constituted by the – globally inconsistent – finite-verb reading of the ambiguous adverbial elements. Thus, the local coherence of the embedded sequence should be dependent on how strong the particular ambiguous word form is associated or linked with the function that constitutes the local coherence. The mental representation of the link between a particular word form and its various meanings is most likely a result of linguistic experience[9]: How often we encounter the word «überrascht» as a finite verb, in relation to its overall frequency of occurrence. In other words, how probable it is that the word form occurs in a context where it is used as a finite verb. Let's assume that corpora can provide a – however rough and incomplete – approximation of the distributional properties of the linguistic environment an "average" language user is confronted with. The hypothesis regarding the visual-world experiment can then be termed as: The higher the probability of the "local" function, given a particular word form, the stronger the local coherence of the embedded sequence. As a consequence, the influence of a locally syntactic coherent substring interferes more strongly with the globally correct interpretation, and will thus lead to a higher proportion of fixations on the local scene.

To test this hypothesis, the *finiteness bias FB* (3.2) was calculated for each ambiguous element contained in the stimuli from experiment 2 (section 3.4), as the tendency of *w* to occur as a finite verb (the locally coherent reading), compared to the overall frequency of *w*:

$$FB(w) = \frac{frequency(w_{FiniteVerb})}{frequency(w)} \tag{3.2}$$

As a first try to evaluate the hypothesis, raw $frequency(w)$ and $frequency(w_{FiniteVerb})$ where derived from a version of the Frankfurter Rundschau (*FR*) corpus, automatically annotated with part-of-speech (POS) information. The corpus consists of newspaper texts of the *Frankfurter Rundschau* from the years 1997-1999 and comprises ~30 million tokens (~75000 types).

3.5.2 Results

3.5.2.1 Influence of Finiteness Bias on Local Coherence Processing

To evaluate whether *FB* shows an effect on local coherence interpretation at all, a subset of the data was analyzed, containing only sentences with local syntactic coherence (condition *+amb/-add.adv (LSC)*, example (43a)). I will refer to this subset as $data_{LSC}$ in the following sections. As in section 3.4.3.2, the period of interest was defined as a time-window spanning from the offset of the adverbial element, plus 200 ms, to the end of the LSC at the offset of the prepositional phrase, plus 200 ms (*loco* in table 3.10). Linear mixed effect models where fitted

[9] I suppose that this claim is relatively independent of any particular account of how linguistic knowledge is represented in the human mind.

to the data with logit of fixation proportions on the local scene as the dependent variable. *FB* was added as a linear predictor, as well as the additional predictors identified in 3.4.3.2. *FB* was added as random slope in both participant and items term. Although the model revealed a numerical tendency in the expected direction of considerable size, compared to most of the other predictors, the effect was not reliable ($t = 1.162$). In figure 3.13, the beta coefficients and the corresponding confidence intervals of the different predictors are plotted. Points located above the dashed line signify a positive effect of the predictor, while points below the line signify a negative effect of the predictor on logit fixation proportions on the local scene. In contrast to the other predictors, *FB* shows a strong positive effect, but also very high variance. Since the missing reliability of the effect seems to mainly result from the high variance, it is reasonable to consider potential reasons for the high variance and try to find ways to eliminate these sources of noise. Two factors can be identified as potential candidates, one concerning the predictor (1), the other one concerning the dependent variable (2).

1. Some of the adverbial participles occurred only very rarely in the corpus. Low frequency of occurrence, however, can heavily skew the finiteness quotient.

2. Fixation proportions are calculated from the summed fixation time on the interest area during the period of interest. This includes trials where participants fixate only the interest area, do not fixate the interest area at all, and trials where saccades into and out of (or both) the interest area occur. Since *FB* differences between items are likely to be rather subtle, it is possible that the effect is not measurable if all types of trials are taken into account. In other words, it is possible that *FB* does not considerably increase the number of trials where only the local scene is fixated, or decrease the number where it is not fixated at all. It may well be that the time spent on the interest area is a more sensitive measure for subtle differences in visual attention.

To overcome the problem of data sparsity due to the limited size of the corpus used, frequencies and corresponding part-of-speech tags were extracted from the bigger *dlexdb* database (www.dlexdb.de; Heister et al. (2011)). The corpora underlying *dlexdb* compass about 100 million tokens and 2.3 million types and thus should provide more reliable *FB* values. In contrast to the *FR* corpus, where 3 of the items occurred only 6 times in the corpus, in *dlexdb* the minimal number of hits was 74 for one item.

However, the analysis on the basis of FB_{DLEX} derived from the considerably bigger corpus underlying *dlexdb* did not even reveal a numerical effect, and exhibits considerable variance (figure 3.13), right panel).

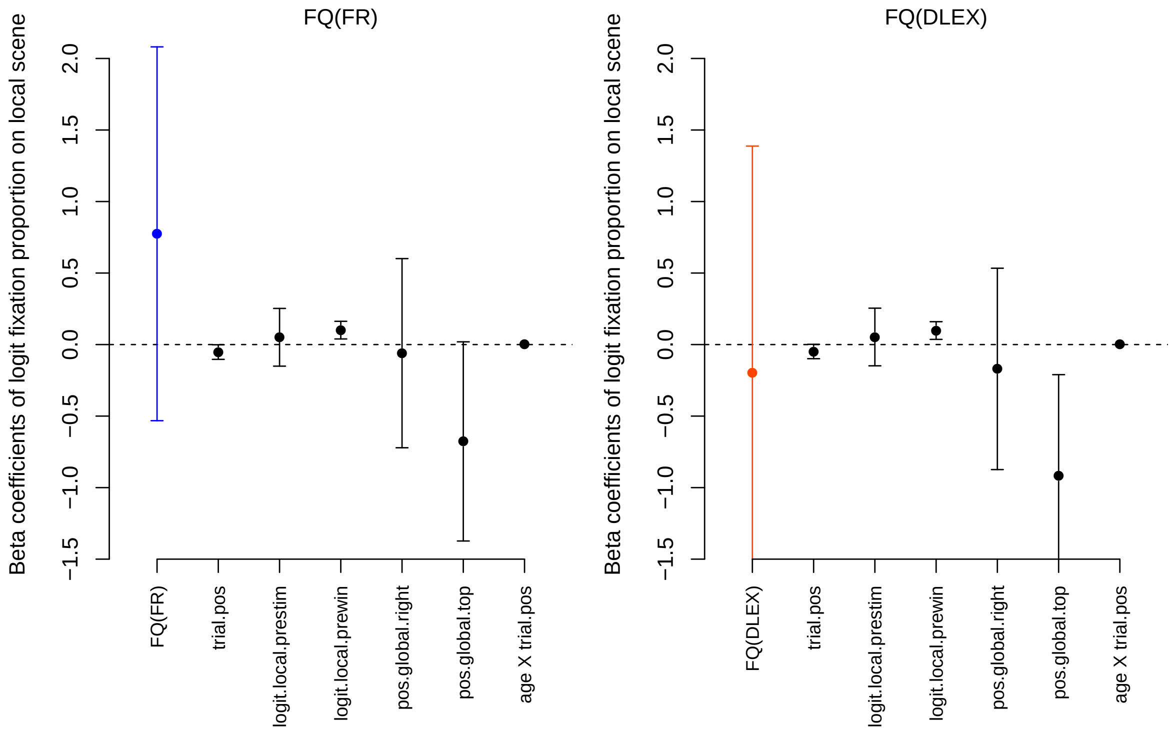


Figure 3.13: Beta coefficients of predictors from LMEM, comparing FB(FR) and FB(DLEX) as predictors of fixation proportions on local scene in sentences with LSC. Only correctly answered trials.

.

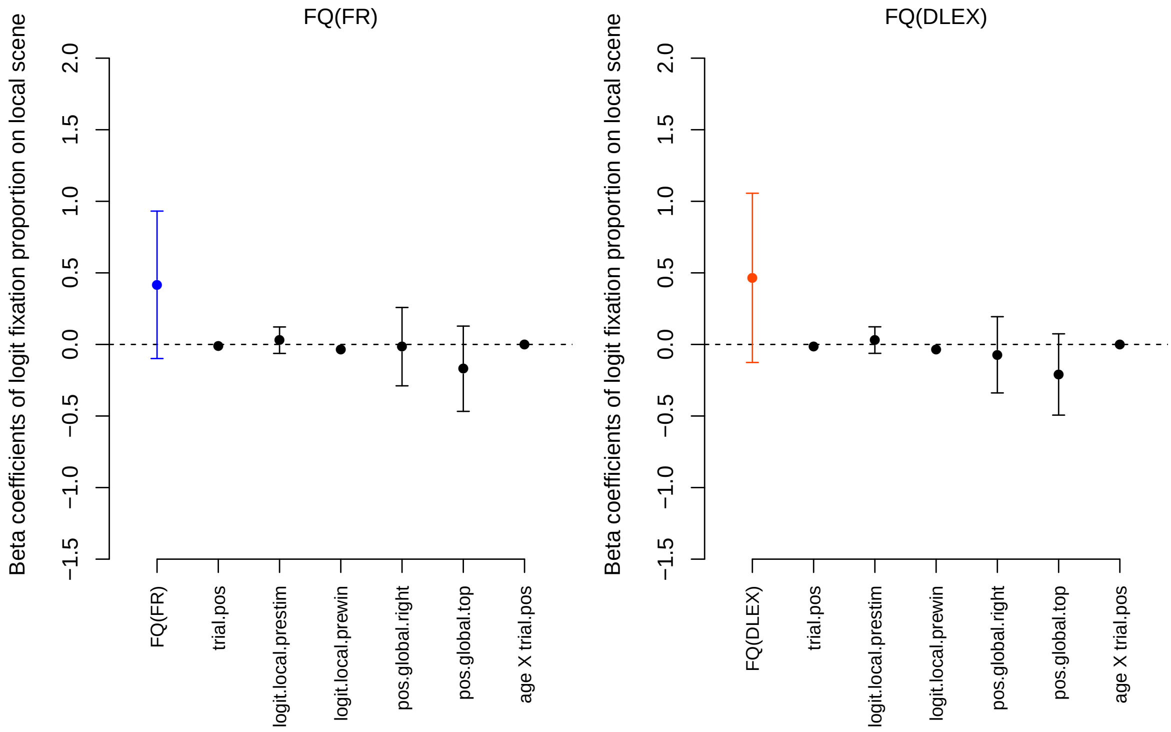


Figure 3.14: Beta coefficients of predictors from LMEM, comparing FB(FR) and FB(DLEX) as predictors of fixation proportions on local scene in sentences with LSC. Only correctly answered trials. Trials with no fixations or only fixations on local scene are excluded.

Excluding trials Given the very high variance for both FB_{FR} and FB_{DLEX}, all trials were excluded where participants during the period of interest either fixated only the local scene (*all*-trials), or didn't fixate the local scene at all (*nothing*-trials), which affected about 33% of the trials. Fitting models with identical structure as described above to the subset of the data revealed, for both corpora, considerably lower variance and numerical tendencies in the same direction. However, for both counts the effect did not prove significant (FB_{FR}: $Estimate = 0.4164$, $SE = 0.2628$, $t = 1.585$; FB_{DLEX}: $Estimate = 0.4650$, $SE = 0.3016$, $t = 1.542$).

Accuracy of part-of-speech tagging However, although *dlexdb* provides a database of considerable size, part-of-speech tagging is done automatically. Since the crucial property of the adverbial elements is their part-of-speech ambiguity, it is crucial that disambiguation is correct. In addition, it is likely that incorrect tag assignment is not equally distributed over cases, but rather that tagging algorithms make systematic errors in particular contexts. *Dlexdb*, however, does not allow access to the underlying corpora, so there is no easy way to check the accuracy of part-of-speech tagging for ambiguous elements. Systematic errors in part-of-speech assignment, however, are a likely cause of considerable noise in our analysis. A look at the tagged parts of the available corpora provided through the COSMAS II interface by the IDS Mannheim (https://cosmas2.ids-mannheim.de/cosmas2-web/) (tagged$_C$, tagged$_T$ and tagged$_M$) confirmed this suspicion: The three subcorpora are annotated with different taggers, but all of them showed only very poor reliability in disambiguating the POS-tags correctly for the ambiguous elements used in experiment 2. Therefore it seemed crucial to manually check and correct the part-of-speech tags to provide a reliable estimate of *FB* for the used stimulus materials. All part-of-speech tags of the adverbial participles in the FR corpus were manually checked and corrected if necessary. It turned out that about 14% of the hits were tagged incorrectly. As mentioned above, some of the the elements occurred only very rarely in the FR corpus. For all elements that occurred less than 30 times in the FR corpus, a random sample of 100 occurrences was drawn from the COSMAS corpus, and part-of-speech tags (finite verb yes/no) were annotated manually. $FB_{corrected}$ was calculated from the resulting set. $FB_{corrected}$ is thus based on at least 30 occurrences of each element, with manually annotated occurrences as finite verbs.

Using $FB_{corrected}$ as a predictor, two models identical in structure as described above were fitted to the $data_{LSC}$, where one included all trials, the other one included only the subset of trials without *all* or *nothing* trials. Numerically, the effect of $FB_{corrected}$ points in the same direction for both datasets. However, while the dataset including all trials shows, as in the analyses before, very high variance, using the smaller subset, a significant effect of $FB_{corrected}$ (see figure 3.15) was revealed.

In the subset, additional predictors, except fixation proportions on the preceding time window (*logit.local.prewin*) did not show reliable effects. Position of the global scene, although not significant, exhibits a relatively strong effect and was therefore also included. All other additional predictors where not included in the model. The resulting model reveals a significant effect of $FB_{corrected}$ ($Estimate = 0.9384$, $SE = 0.3750$, $t = 2.503$) such that participants spent higher proportions of fixation on the local scene for ambiguous adverbial elements with higher $FB_{corrected}$, that is, if an element is more likely to occur as a finite verb. The simpler model is summarized in table 3.11.

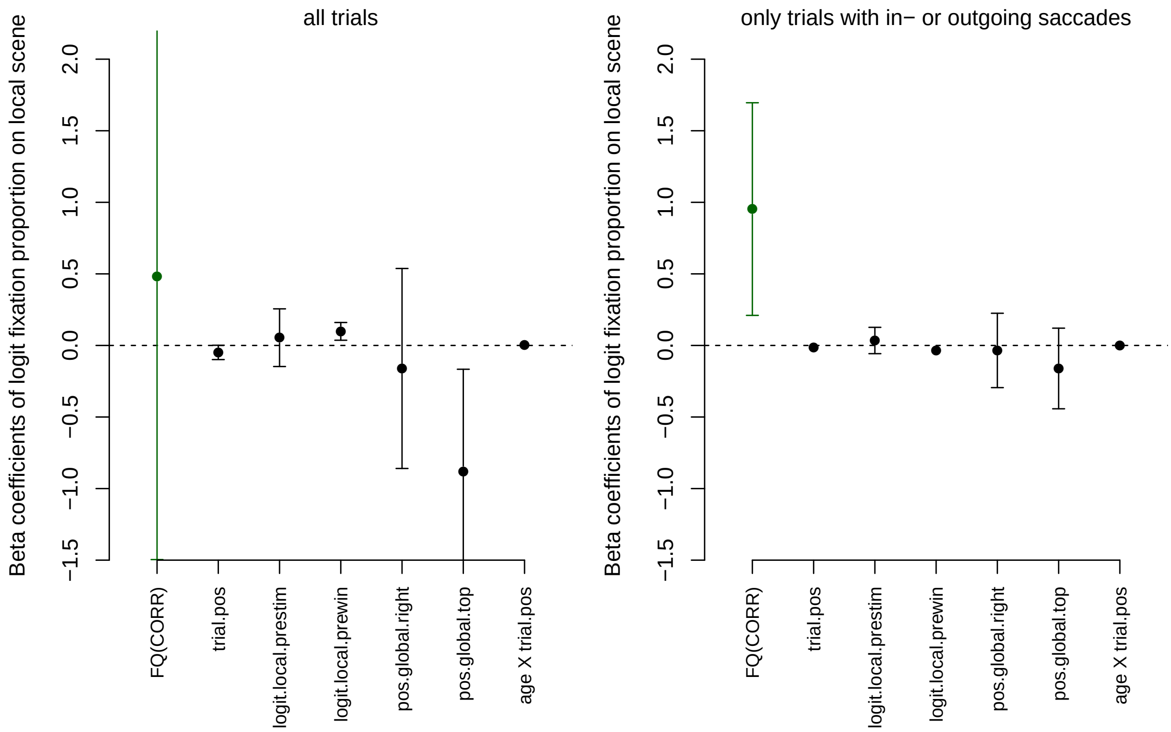


Figure 3.15: Beta coefficients of predictors from LMEMs. FR(CORR) is included as linear predictor of fixation proportions on local scene in stimuli with LSC. Left hand panel: all correctly answered Trials. Right hand panel: trials with no fixations or only fixations on local scene are excluded. Whiskers indicate confidence intervals.

Table 3.11: Models of logit of fixation proportions on local scene, condition *+amb/-add.adv (LSC)*; only correctly answered trials. Trials without fixations on the local scene as well as trials with fixations only on the local scene are excluded from the analysis.

Factor	Estimate	Std. Error	*t*-value
Intercept	-0.8488	0.1295	-6.557
FB_{CORR}	0.9385	0.3750	2.503
logit.local.prewin	-0.0321	0.0139	-2.304
pos.global.right	-0.0337	0.1325	-0.255
pos.global.top	-0.1666	0.1433	-1.163

model formula (R syntax): logit.local.win (LSC) ~FB_{CORR}
+ logit.local.prewin + pos.global +
(1 + FB_{CORR} | item) + (1 + FB_{CORR} | participant)

3.5.2.2 Ruling out Context Independency

So far only sentences with local syntactic coherence were included in the analysis of *FB* influence. The results regarding gradient effects in local syntactic coherence processing up to this point provide evidence that the strength of local coherence, estimated from corpus data, has a significant effect on gaze behavior that mirror interpretational processes during listening to the sentence. However, since *FB* only measures properties of the ambiguous element, we can not readily infer from these results that the effect provides evidence for local coherence processing conceived as depending on local syntactic context. The influence of *FB* may just reflect retrieval of different senses of the ambiguous word while listening, regardless of the surrounding context. Given that the finite verb meaning of an element matches with the local scene, participants' eye movements may be attracted by this scene, regardless of any syntactic properties of the context, and regardless of the argument structure the element is embedded in, and would thus be compatible, for example, with Gibson's (2006) proposal of interference of context-independent lexical category frequency. Gibson only refers to the self-paced reading data from Tabor et al. (2004) and thus mainly posits an influence on reading time, but remains silent about the potential influence of these lexical properties on interpretational processes. However, the proposal at least does not explicitly rule out that interpretational processes may also be triggered by context-independent bottom-up effects. Thus, to rule out that the effect is exclusively due to context-independent lexical effects, it is necessary to show that the preceding context interacts with the lexical properties of the ambiguous elements.

With regard to the experimental design of the visual world experiments, this can be termed in the following hypothesis: If higher *FB* triggers eye movements to the local scene regardless of the preceding context, the effect of *FB* on fixation proportion should not only show up in sentences with local syntactic coherence, but also in sentences that contain the identical ambiguous element in a different context. This is the case in *+amb/+add.adv* example (43c), which contains an ambiguous adverbial element, that, however, is preceded by the additional adverb «äußerst», which inhibits local coherence by rendering the resulting substring ungrammatical. If, in contrast, the effect of *FB* is eliminated by the additional adverb, purely context-independent processes are rendered rather unlikely.

To disentangle these hypotheses, in the next section I will analyze the subset of the data including only stimuli with local syntactic coherence (condition *+amb/-add.adv (LSC)*, example (43a)) and stimuli where local syntactic coherence is inhibited by the additional adverb (condition *+amb/+add.adv*, example (43c)). Since all stimuli in this subset contain an ambiguous element, I will refer to it as $data_{AMB}$ in the remainder of the chapter.

Results Linear mixed effect models where fitted to the described subset $data_{AMB}$. In the preceding section it turned out that an effect of *FB* is only reliably observed when excluding trials where participants fixated only the local scene or did not fixate it at all during the period of interest. Therefore, the trials were also excluded in the current analysis. Again, only correctly answered trials were included. Given that FB_{CORR} provides the most reliable measurement with regard to part-of-speech accuracy, FB_{CORR} was added as linear predictor. Presence or absence of an additional adverb *add.adv* was centered to the mean of the binary coding of the factor (with *-add.adv* represented as positive value). Both factors and their interaction are included as linear predictors, and were also included as random slopes in participant and item terms, providing the maximal random effect structure. Of the additional factors, only fixation proportions on the local scene in the preceding time window (*logit.local.prewin*) and fixation proportions on the local scene before the onset of the stimulus sound (*logit.local.prestim*) showed reliable effects. All other additional factors were not included in the model.

The resulting model did not reveal significant main effects of the factors, and no significant interaction (table 3.12). However, the numerical values show the expected pattern such that fixation proportions on the local scene increase with higher FB_{CORR}, but only in absence of the additional adverb (figure 3.16). Accordingly, fitting models only to the subset of the data with the the target condition *+amb/-add.adv (LSC)* revealed a significant effect of FB_{CORR} ($Estimate = 1.0086$, $SE = 0.3729$, $t = 2.705$). FB_{CORR}, however, had no effect when analyzing only the control condition *+amb/+add.adv* with the additional adverb ($Estimate = 0.1614$, $SE = 0.5012$, $t = 0.322$) (figure 3.17).

Table 3.12: Models of logit of fixation proportions on local scene, comparing target condition *+amb/-add.adv (LSC)* (example (43a)) and control condition *+amb/+add.adv* (example (43c)) with additional adverb; only correctly answered trials. Trials without fixations on the local scene as well as trials with fixations only on the local scene are excluded from the analysis.

Factor	Estimate	Std. Error	t value
Intercept	-0.8345	0.0938	-8.899
FB_{CORR}	0.5830	0.3502	1.665
add.adv	-0.1451	0.1368	-1.061
logit.local.prestim	0.0627	0.0330	1.898
logit.local.prewin	-0.0484	0.0109	-4.443
$FB_{CORR} \times$ add.adv	0.8545	0.5880	1.453

model formula (R syntax): logit.local.win ~FB_{CORR} * add.adv
+ logit.local.prestim + logit.local.prewin +
(1 + FB_{CORR} * add.adv | item) + (1 + FB_{CORR} * add.adv | participant)

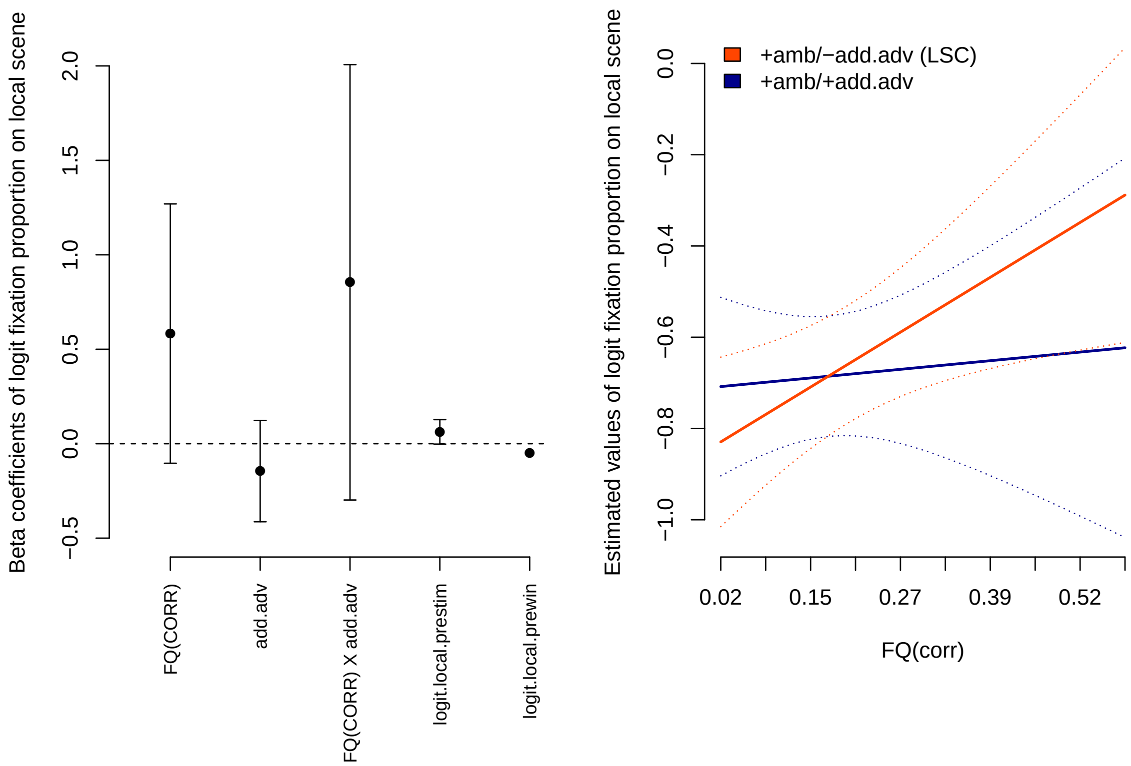


Figure 3.16: Beta coefficients (left) and estimated values (right) from LMEMs, including stimuli with LSC and control condition with additional adverb. Trials with no fixations or only fixations on local scene are excluded. Whiskers/dotted lines indicate confidence intervals.

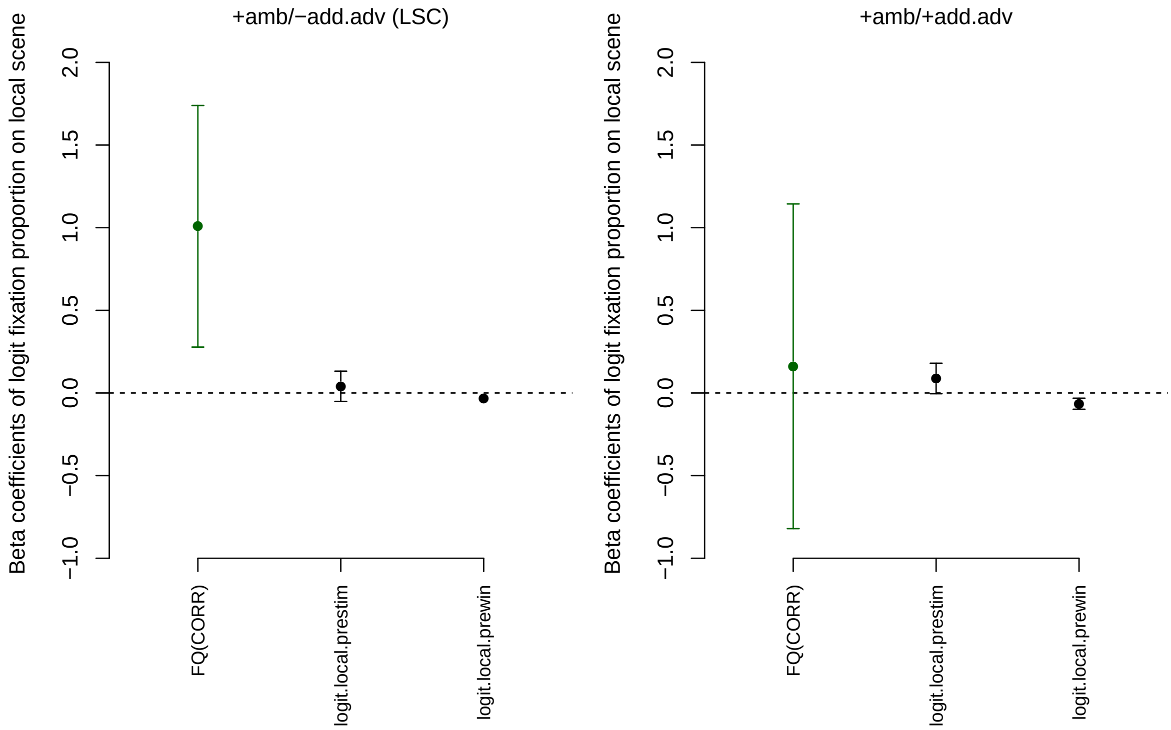


Figure 3.17: Beta coefficients from LMEMs. Left hand side: stimuli with LSC. Right hand side: control condition with additional adverb. Trials with no fixations or only fixations on local scene are excluded. Whiskers indicate confidence intervals.

3.6 Discussion

In the following section, I will first discuss two potential issues with regard to the design and the materials used in the two visual world experiments. Following that, I will discuss the results with regard to the approaches that have been put forward to account for effects of local syntactic coherence, or that have been identified as being potential candidates in section 2.3.

3.6.1 Critical Issues

3.6.1.1 Influence of Visual Depictions on Language Processing

In the visual world experiments participants can inspect the scenes before they hear the sentence. Thus, it might be argued that the presence of a scene depicting the locally coherent meaning is the main reason why participants fixated it. Roughly sketched, participants build up possible descriptions of the three scenes while inspecting them, and as soon as the sentence is played, compare these descriptions with the auditive input, finally fixating the scene where the prebuilt descriptions best match the stimulus sentence heard. This matching process would then be considerably different in cases where people only listen to a sentence, and cannot select between prebuilt propositions.

However, to access a potentially already built proposition, the locally coherent sequence still has to be noticed while listening to the sentence. It cannot be ruled out completely, however, that the visual presentation of the local meaning enhances the likelihood that the locally coherent sequence is processed, or how deep this processing will be.

The combination of visual and linguistic input, however, always bears the potential problem that it is hard to know if the processes can be generalized to language processing in situations without visual input. In short, it is very likely that eye movements do not exclusively reflect processes of language processing, but that the available additional visual information is actively used to understand the sentence, and that the processes differ considerably from processing language in other situations (c.f. Huettig et al., 2011). Although it is rather uncontroversial that the visual world paradigm provides a very sensitive tool to investigate which sources of information *can* be used during processing, it does not necessarily prove that all of these sources are also used in other scenarios of language comprehension, where no additional and potentially helpful – or, with regard to the experiments reported here, distractive – visual input is present.

3.6.1.2 Splicing and Prosodic Cues

Another issue is related to the preparation of auditive stimuli. As described in 3.3.1, a large part of the locally coherent sequences was taken from a separately recorded main clause. The underlying idea was that prosodic cues may shadow the potentially subtle effect of local syntactic coherence, and thus the intention of the manipulation was to eliminate this potential disturbance. Therefore, it may well be that the effects would have been weaker (or even invisible) without that manipulation.

The critical argument is that the effects are *solely due to* this manipulation, such that the main clause prosody induced interpretational processes that would not have been triggered otherwise. The results cannot completely rule out the latter hypothesis. However, even if the latter hypothesis holds true, it is still remarkable that such a subtle prosodic manipulation which is hardly noticed at all is able to partially overwrite the rather strong syntactical constraints that should rule out consideration of the local meaning right away.

In conclusion, the slicing issue provides a potential weakness of our results. To rule out the hypothesis that the effects are mainly induced by latent prosodic cues and to provide stronger evidence that effects of local syntactic coherence can be conceived as a phenomenon that mirrors fundamental mechanisms in human sentence processing, the next steps will have to provide evidence for the effects without prosodic manipulation of the stimuli.

3.6.2 Possible Explanations

In the preceding sections, we provided further evidence for the influence of local syntactic coherence on sentence processing. The results show that these sequences do not only provide a source of complexity on the structural levels, but that the local sequences are interpreted to a certain degree, resulting in higher fixation proportions on scenes that represent the meaning of the local sequence, reflecting that some amount of attentional resources is assigned to the "ungrammatical" local parse. In addition, we provided evidence that the amount of attention is strongly modulated by the distributional properties of the local sequence. The more likely the ambiguous element occurs in a function that supports the local interpretation, the greater the amount of attentional resources that is dedicated to the respective scene. Given these results, we will now review the attempts described in section 2.3, that have been put forward to account for earlier evidence of local syntactic coherence, and discuss if and in which way they can account for our results.

As noted above, the presented data can hardly single out one particular model as superior to all the others. In general, most of the approaches explicitly addressed effects of local syntactic coherence on reading time and thus do not make explicit claims about eye movements in the visual world paradigm. However, I will go through the proposed models in turn and focus mainly on two questions: Firstly, would the approach predict any differences in processing that resemble the reported results? And, secondly, are there good reasons that, based on the underlying assumptions of the approach, these differences should be mirrored in fixation patterns in the visual-world paradigm?

3.6.2.1 SOPARSE and Impulse Processing

Since the experiments are closely related to the design used in Tabor et al. (2004), it is not surprising that the SOPARSE model would predict an effect of local syntactic coherence in our experiment as well. The ambiguous element results in the activation of a lexically anchored tree fragment that competes to be linked with the preceding noun phrase. However, the model does not readily provide a mechanism to model interpretation and is rather construed to model the (self-paced) task of reading and not to the externally paced listening process. However, the underlying framework of dynamical systems, which is explicitly applied to visual-world data in Kukona and Tabor (2011), provides a plausible characterization of the process: The locally coherent sequence provides bottom-up impulses that change the attractor landscape of the system such that the attractiveness of the scene depicting the local meaning grows larger than in the control sentences. Thus, given the stochastic selection process of fixated regions, the probability of fixating the local scene is increased. Crucially, the approach does not require an explicit representation of the local content. Rather, the bottom-up information can be understood as disturbing the system such that, in case of a self paced process, it needs more time to settle in a stable state, or, in case of an externally paced task as in the visual world paradigm, is drawn away from the globally correct attractor towards the local scene, resulting in a higher likelihood of fixating it. Regarding interpretation, under this perspective it is more adequate to conceive the system state not as representing two meanings in parallel, but rather as representing an *intermediate* meaning in between local and global interpretation.

3.6.2.2 Word-Based Approaches

As has already been discussed with regard to experiment one, Gibson's (2006) approach of bottom-up relative category frequency can in principle account for the difference between locally coherent stimuli and controls with unambiguous adverbs. In addition, the influence of finiteness bias, although operationalized slightly differently, is consistent with Gibson's approach. However, the fact that both the difference and the influence of finiteness bias only appear if local syntactic coherence is present, but not when an additional adverb inhibits local coherence, is clearly conflicting with that view. In particular, while in the results from Tabor et al. (2004), the difference between reduced and unreduced relative clauses provides rather different syntactic expectations and, as Gibson attempts to show, result in different results of the interaction between syntactic top-down and lexical bottom-up components, this is not true for the stimuli used here. Although the preceding word differs, it is unlikely that it changes top-down prediction of syntactic category considerably.

Regarding the visual-world data as such, since it is proposed that the effects merely reflect retrieval processes at the lexical level, it is not clear why the interference should result in fixations towards the local scene.

Regarding the cue-based retrieval approach provided in Van Dyke (2007), it is hard to conceive of a clear prediction with regard to our experiments. Since the approach is mainly concerned with interference during the retrieval of verb arguments when the verb is encountered, there seems to be, at least at first, no easy way to account for differences that do not involve argument retrieval. In addition, since the arguments are identical in all conditions, no similarity-based differences can be brought up as an explanation for the observed results. However, a tentative explanation could be that, since the relative clause requires a verb, on encountering an ambiguous element that can occur as a verb, the corresponding lexical entry is retrieved accidentally in a number of cases. If, however, the element is conceived as a verb it will be longing for its arguments to be retrieved, and the most likely candidate will be the preceding noun, the "astronaut", since it matches both in category and case and experienced the least activation decay. Such a process will result in a unwarranted proposition representing the astronaut as the subject of surprising, and, assuming direct triggering of eye movements, will lead to fixations on the local scene. Regarding the control conditions with the additional adverb, the likelihood of retrieving the argument will be decreased just due to the fact that more time has passed since the astronaut has been processed. However, a detailed model of how to incorporate both visual processing and lexical ambiguity within the cue-based parsing approach is neither the aim nor feasible here.

3.6.2.3 Modular Approaches

As has been noted before (section 2.3.4), modular approaches such as Bicknell et al. (2009) and Corley and Crocker (2000), as well as the sausage machine (Frazier & Fodor, 1978), can in principle account for effects of local syntactic coherence – at least in terms of processing cost. With regard to the influence of finiteness bias, the structural approach taken in the sausage machine, can not really be accounted for. In contrast, the statistical n-gram based approaches would obviously also predict the observed influence of the relative category distribution on the effects. However, the explicitly stated modularity of the proposed architectures of Frazier and Fodor (1978) and Corley and Crocker (2000) make it difficult to account for the fact that participants fixate a scene that represents the meaning of output of the proposed module. Since the motivation of a modular approach is efficiency, a crucial property of a module is that the representational format of both input and output of the module is relatively restricted (Crocker & Corley, 2002, p. 3). Thus, the module itself will not generate hypotheses about the meaning of a substring that could trigger eye movements to a corresponding scene. In particular, as far as I understand, the purpose of the modular mechanisms is to provide the next stage of processing with a fast first guess about the structure of partial sequence. However, this first guess is then immediately compared with the global structure that has been built up in order to arrive at a consistent interpretation, and work has to be done if this first guess does not fit, resulting in the observed pattern on an abstract level. However, within this picture it is hard to incorporate that such a first guess, although completely incompatible with the global structure, should be allowed to be processed any further.

In other words, it is inconclusive to provide a mechanism to ensure self-consistency, or, for that matter, rational strategies on the syntactic level, but, metaphorically speaking, to release the intermediate products of the preprocessing module to consume valuable resources by triggering eye movements. However, since none of the approaches has been explicitly set up to account for neither spoken language nor its interaction with visual input, it may well be that an adequate linking hypothesis could be formulated that may close this gap.[10]

[10] The predictions of the respective models, however, may differ depending on the particular specification of the window-size and, crucially, the informational sources and the grain-size the preprocessing is operating on. As Tabor et al. (2004) point out, to account for the semantical effects in their experiment 2, at least animacy information has to be taken into account. However, since we did not include this factor in our design, the results do not readily suggest any particular claim with regard to this matter.

3.6.2.4 Rational Parser

Hale's *rational parser* (2011) explicitly includes effects of local syntactic coherence as a main motivation for his attempt, and explicitly models the observed pattern with regard to the stimuli of experiment 1. Since the design of experiment 2, as far as I can see, does not differ with regard to relevant properties regarding the model, it is reasonable that the parser similarly visits more intermediate states in case of local syntactic coherence. It is, however, not entirely clear how these intermediate states relate to the observed fixation patterns. However, as Hale shows for the results from Tabor et al. (2004), a number of intermediate states include the locally coherent partial parse tree as a main clause structure. Given the additional assumption that eye movements are partly automatically triggered by those intermediate states, based, for example, on a stochastic process, the observed pattern should emerge. However, similar to the other accounts that do not directly address the further results of a (temporal or final) syntactic representation of the input, the link between constructed parses and eye movements remains unclear.

3.6.2.5 Noisy Channel

The noisy-channel model (Levy, 2008b; Levy et al., 2009) described in 2.3.5.2 provides a potential, alternative explanation for at least part of the observed effects. Considering experiment 2, in the locally induced proposition «die Astronautin überrascht den Außerirdischen» the astronaut is the agent of «überraschen» and the alien is the patient. Let's consider the idea implemented in the noisy-channel model that the language processing system may reconsider the already processed input under particular circumstances. Within the dative relative clause, the adverbial element «überrascht» is not obligatory, and is a rather unexpected constituent at that position. Given that the particular ambiguous element is more likely to be encountered as a finite verb than as an adverbial participle in general, a comprehender may, instead of trying to integrate the word as an adverbial participle, reconsider the hypothesis about the already heard input, figuratively asking the question: Is there a good chance that a slight change in the input processed so far, would give me a more likely syntactic structure confronted with the current input word? Actually, there is a rather good chance: The dative case marking of the relative pronoun «deM» is the relevant factor that renders the finite-verb interpretation of «überrascht» ungrammatical, giving its preceding context. A near neighbor that differs in only one letter (or, for that matter, one phoneme) is the accusative relative pronoun «deN», which would render the finite-verb reading perfectly grammatical, and would even suggest a less complex syntactic structure.

(44) Hugo, deN die Astronautin überrascht ...
Hugo, who$_{acc}$ the astronaut suprises ...
Hugo who the astronaut surprises ...

Example (44) would constitute a accusative object relative clause with the astronaut surprising somebody. In contrast to the local syntactic coherence, however, it is then «Hugo» that is getting surprised, but not the alien. Although no scene in the visual display corresponds to this meaning, the interpretation still constitutes the astronaut as surprising someone – in contrast to doing something surprisedly, and may thus trigger eye movements towards the local scene where this is the case. Since the unambiguous control «ungläubig» does not provide a similarly good reason to reconsider the input, this could result in the observed difference in fixation proportions after having heard the adverbial element.[11]

[11] Another possibility would be to change «deM» to the nominative relative pronoun «deR», leading to a subject relative clause:

(1) Hugo, deR die Astronautin überrascht ...
Hugo, who$_{nom}$ the astronaut suprises ...
"Hugo who surprises the astronaut... "

The subject relative clause would represent Hugo as the agent and the astronaut as the patient. However, this is even less compatible with the visual input, and would hardly explain fixations on the local scene where the astronaut surprises somebody. However, things get complicated as soon as the sentence continues. In 7 of the 48 stimuli, as in the current example, the hypothesis that one phoneme has been misunderstood is not disproved within the entire sentence, since the second noun phrase of the relative clause «den Außerirdischen» is ambiguous between accusative singular and dative plural, leading to the coherent sentence:

(2) Hugo, den die Astronautin überrascht den Außerirdischen vom Mars zeigt, ...
Hugo, that$_{ACC}$ the astronout$_{NOM}$ surprisedly the aliens from Mars shows, ...
"Hugo that the astronaut shows suprisedly to the aliens from mars ... "

However, in the potentially hypothesized sentence, expressing that the astronaut shows Hugo surprisedly to a group a aliens, «überrascht» is, again, not regarded as a finite verb but as an adverbial participle, just as in the original dative relative clause. Again representing the astronaut as *showing* something (and not *surprising*), but Hugo and the aliens have switched roles. Thus, fixations may be triggered, again, to the global scene (which, finally, is the case), that, however, does not readily fit the proposition. An interesting point is that a comprehender, by revising his beliefs about previous input for good reasons, will be led down a garden-path that actually did not exist at the beginning, and will thus again have to revise the interpretation.

Of course, it may also be the case that, as soon as «the astronaut» is heard, the visual information is used to rule out the temporally reconsidered accusative ORC hypothesis and the globally correct hypothesis will win out.

Since the noisy-channel model does not make any direct claims about eye movements in the visual world paradigm, but rather on processing costs mirrored in reading-times, these speculations are rather tentative. However, we can assume that eye movements are triggered to fixate aspects in the visual environment that fit with the most plausible hypothesis about the sentence meaning (or aspects that may be helpful to distinguish between these hypothesis). Given our example, participants thus may attend to the scene depicting the astronaut as agent as soon as they hear «überrascht», but then switch back to the globally consistent interpretation. The data of our experiments, however, do not allow for a detailed evaluation of this or a similar hypothesis.

While, as has been discussed earlier, an explanation solely based on lexical category ambiguity would predict the same difference regardless of the additional adverb («gerade» and «äußerst» in experiments 1 and 2, respectively) (see section 3.6.2.2), this is different with regard to the reconsideration hypothesis. In experiment 2, «äußerst» was used to avoid a sequence that resemble the beginning of a sentence: «Äußerst» cannot be used to modify a finite verb. However, this property would also lower the chance that the ambiguous element initiates reconsideration of the initial hypothesis about the sentence structure, because changing «deM» to «deN» does not result in a grammatical structure if «äußerst» precedes the ambiguous element:

(45) * Hugo, deN die astronautin äußerst überrascht, …
 * Hugo that$_{acc}$ the astronaut very surprises, …

Thus, the difference between ambiguous and unambiguous items including «äußerst» is not readily predicted, which would result in the observed pattern.

In contrast to assuming that comprehenders consider a misunderstanding, it may of course also be the case that they actually *do* misunderstand the relative pronoun, at least in some cases. If so, the ambiguous element again constitutes a garden path in the locally coherent condition. Given the likely assumption that the finite verb interpretation is both easier (in terms of, for example, minimal attachment) and more likely (in particular for words in the cases with high finiteness bias), participants will initially go for the finite verb interpretation

However, in all other stimuli, the adapted finite verb reading is disambiguated at the case-marked offset of the noun, since the form is not ambiguous between accusative singular and dative plural.

(3) Dieter, dem die Schwimmerin amüsiert den Surfer am Strand präsentiert, …

(4) Ullrich, dem der Barkeeper belustigt die Tänzerin aus Ulm ankündigt …

representing the astronaut as surprising Hugo, and will search for a scene that matches this proposition, which would result in a similar pattern as described above

This explanation is similar to the garden-path by intransitive reading issue already discussed in 3.3.4. However, while the intransitive reading of the used verbs in experiment 1 is rather uncommon, the ORC reading in experiment 2 requires the common transitive usage and thus may invoke a stronger effect. As with the reconsideration hypothesis, the use of «äußerst» instead of «gerade» would not inhibit the interpretation of the ambiguous element as the relative clause verb. Thus, misunderstanding of the relative pronoun could be a possible explanation of the observed pattern, and would also be consistent with the missing influence of the finiteness bias in conditions with an ambiguous element and the additional adverb. However, both misunderstanding and the reconsideration hypothesis can hardly account for the data in experiment 1, unless we assume that participants understand or exchange the complementizer by an accusative relative pronoun.

The hypothesis that can be derived in terms of the noisy channel approach is that participants may reconsider the relative pronoun in face of the ambiguous element and thus maneuver themselves into a garden-path that is actually not present in the input. At least on the basis of the intuitive consideration of the argument given here, such an explanation may account for some aspects of the observed data, although it does not readily account for the observed pattern as a whole. However, it remains to be seen if more comprehensive modeling of the stimuli under a noisy channel perspective could do so. A potential question regarding this explanation, however, is if the proposed strategy is as "rational" as it is construed, if it leads to problems that were not there in the first place.

3.6.2.6 Good Enough Parsing

Somewhat similar to the modular approaches discussed above, the good-enough approach argues for two different pathways that are used to build up an interpretation of a sentence. On the one hand, a syntactic pathway uses morphosyntactic information and knowledge. On the other hand, the semantic pathway uses semantic knowledge (who typically does what to whom, what word order normally signifies which argument structure), owing to Bever (1970) and his proposal of perceptual strategies, including the *N-V-N* template. As has been mentioned above, the "good-enough" parsing approach can in principle explain the results and, in contrast to modular approaches, could also explain interpretational processes.

With the experiments presented here, we are not able to readily rule out any of the attempts, although Gibson's (2006) account seems rather unlikely to be able to account for the data without major changes in fundamental assumptions. Several models can account for the general pattern observed on the two experiments, as the modular approaches, the noisy channel model or Hale's rational parser. However, none of them provides an explicit hypothesis regarding the link between syntactic and further processing. However, for the most part, it is rather unlikely that actual behavioral responses operating on locally coherent structures can be incorporated into the models without stipulating processes that are either not consistent with, for example, the proposed modular architecture or the basic motivation underlying the approaches to provide a *rational* model of sentence comprehension. The focus on processing costs and, accordingly, on reading data, does not allow for a direct link to our data.

Another point worth mentioning is that the use of the visual world paradigm, as has been pointed out above, makes it necessary to consider the likely possibility that eye movements in the visual world paradigm do not only reflect interpretational processes, but rather that visual information is used as an additional source of information to interpretation. The effects can thus both be viewed as evidence that local syntactic coherence is processed and interpreted, and/or that cues in the visual environment can not only support interpretation, but can also interfere with the syntactic processing. The latter point could be incorporated within the noisy channel approach by inducing visual information as an additional factor that may lead to reconsidering the hypotheses with regard to the already processed linguistic input. If both distributional knowledge and visual information make an interpretation likely that does not match with the (assumed) structure of the input processed so far, it may be a good idea to explore more or equally plausible paths, possibly including ones that require a considerable change with regard to the hypothesis so far.

Dynamical system approaches can, in my view, best account for the data. On the one hand, they inherently propose a highly interactive and parallel processing mechanism and can thus incorporate the dynamical integration of not only different linguistic levels, as in interactive constraint-based approaches, but also, as has been shown in Kukona and Tabor (2011), the integration and interaction of multiple sources of sensory input. Crucially, they provide a way of explaining the phenomenon not by categorizing the effects as an unwarranted side-effect of a particularly efficient or rational mechanism, but as an unavoidable outcome of the basic functioning of the system as a whole, which also includes its capability to learn a language in the first place.

4 Modeling Effects of Local Syntactic Coherence with Simple Recurrent Networks

In the preceding chapter, I showed that local syntactic coherence has an effect not only on reading-times, as has been shown in previous research, but also on eye movements in the visual-world paradigm, providing further evidence that locally coherent sequences trigger interpretational processes. In addition, the corpus-based analysis revealed that the size of the effect depends on the strength of the locally coherent sequence, estimated as the finiteness quotient of the ambiguous elements, representing their bias towards the local, finite verb reading. I argued that the results, on the one hand, provide evidence against approaches attempting to explain previous results as effects of context-independent lexical processes (Gibson, 2006), and, on the other hand, support a view on human language processing mechanisms as a highly interactive, self-organizing system, as it is underlies connectionist and dynamical-system approaches to cognition (Tabor et al., 2004; Kukona & Tabor, 2011). From this view, effects of local syntactic coherence arise as an inevitable consequence of the inherent mechanisms that operate on the distributional properties of the linguistic environment without an explicit representation of grammar as a rule-based system, resulting in effects of locally induced, but "ungrammatical" predictions built on frequently encountered patterns in the environment. In particular, this view stands in stark contrast to theories that attempt to explain the effects as mirroring particular, "rational" mechanisms within a rule-based, self-consistent parsing mechanism (Hale, 2011; Levy, 2008b).

In section 2.2.3.1, I introduced simple recurrent network models as implementations of an experience-based perspective on sentence processing that provide a simple, but powerful device of acquisition and processing sequential input. In this section, I will provide a simple recurrent network model to evaluate whether it can serve as a plausible candidate to explain effects of local syntactic coherence as a side-effect of the self-organizing character of acquisition and processing in such a device.

Rather then providing an explicit cognitive model of the visual world data, the SRN model will mainly focus on the underlying mechanisms that, as I will argue, provide the basic foundations of a more comprehensive model that could provide a direct link to behavioral data. Within the SRN framework, effects of local syntactic coherence will be conceived as resulting from *false alarm* activation of nodes in contexts where they are not grammatically licensed, but are strongly predicted by local context, resulting from frequent encounters of the respective local sequences in other contexts. Crucially, these false alarm activations do not hinder the network to successfully predict the correct elements, but rather cause a slight distraction.

The aim of the model is

1. to provide a rather simple mechanism that is able to learn and process a language that resembles human language in some important properties

2. to evaluate if the successfully trained model elicits effects of local coherence as false alarm activations at relevant positions in the input stream, resembling the pattern of results observed in the visual world experiments reported above

To achieve the first aim, it will be necessary to provide training materials that represent a reasonably complex language, and to apply methods to reliably evaluate whether this language has been learnt. Regarding the second point, the materials will have to include structures containing locally coherent sequences, and similar structures that can be reasonably compared. Since local syntactic coherence implies that the embedded sequence also occurs (frequently) in different contexts, it has to be ensured that these other contexts exist within the language.

In the following sections, I will first motivate and describe the materials, architecture and procedure that have been used with the described aims in mind (section 4.1). Thereafter, I will evaluate the performance of the model with regard to the first aim (section 4.2), followed by the investigation of effects of local coherence in section 4.3. A discussion of the model, its relation to empirical results as well as potential weaknesses of the approach adopted here will be given in chapter 5.

4.1 Materials, Architecture and Procedure

4.1.1 Grammar and Training Corpora

The materials to train the networks were generated with a probabilistic context free grammar (*PCFG*). The *Simple Language Generator* (Rohde, 1999) provides a powerful and flexible tool to write such a grammar and generate the desired number of sentences according to this grammar. In addition, the tool calculates the transitional probabilities at every word position for a set of given sentences according to a provided grammar, which is crucial to evaluate network performance.

The grammar is designed to implement a reasonably complex language that roughly mirrors a subset of German. It includes sequences that constitute LSCs of the kind used in the visual-word experiments. In addition, I included different kinds of relative clauses to be able to evaluate if the networks are able to process complex sentences and thus can be taken as plausible models of central aspects of human language comprehension. On the other hand, the grammar is much less complex than real German in several aspects. These simplifications are necessary, on the one hand, to provide a model that is simple enough to be analyzed in detail. Secondly, since the architecture used is a basic SRN and thus a rather simple model without special adaptations, the format of representation, power of the architecture, and, not at least, time constraints advise to keep the grammar as simple as possible while still entailing the relevant structures and variants necessary to investigate the question at hand.

The materials we used in our simulation were generated according to the grammar shown in table 4.1. A full version of the working SLG-grammar can be found in appendix D. I will describe some aspects of the grammar in the following section.

The lexicon consists of 32 elements, including

- 4 transitive verbs, 2 of them ambiguous (transitive verb singular/adverb);
- 2 ditransitive verbs;
- 2 masculine nouns, with determiners;
- relative pronouns (nominative, accusative, dative);
- 2 ambiguous (adverb/transitive verb singular) and 2 unambiguous adverbs;
- 1 modifying adverb;

- end-of-sentence marker (EOS);

Table 4.1: Phrase structure rules of the probabilistic context free grammar used to generate training materials. | signifies alternative expansions; "" signifies an empty element; bracketed numbers signify selection probabilities. Constraints to control number agreement and argument structure are not included. The complete grammar in SLG format can be found in appendix D

1	S	→	NP_{nom} VP EOS
2	NP_{nom}	→	N_{nom} RC
3	VP	→	V indObj ADVP NP_{acc}
4	indObj	→	NP_{dat} | ""
5	V	→	V_{trans} | $V_{ditrans}$
6	NP_{acc}	→	N_{acc} RC
7	NP_{dat}	→	N_{dat} RC
8	RC	→	SRC (0.03) | ORC (0.06) | ""
9	SRC	→	RP_{nom} NP_{acc} ADVP indObj V
10	ORC	→	ORC_{acc} | ORC_{dat}
11	ORC_{acc}	→	RP_{acc} NP_{nom} ADVP NP_{dat} $V_{ditrans}$ | RP_{acc} NP_{nom} ADVP V_{trans}
12	ORC_{dat}	→	RP_{dat} NP_{nom} ADVP NP_{acc} $V_{ditrans}$
13	N_{nom}	→	der-pilot-*NOM-SG* | die-piloten-*NOM-PL* | der-kommandant-*NOM-SG* | die-kommandanten-*NOM-PL*
14	N_{acc}	→	den-piloten-*ACC-SG* | die-piloten-*ACC-PL* | den-kommandanten-*ACC-SG* | die-kommandanten-*ACC-PL*
15	N_{dat}	→	dem-piloten-*DAT-SG* | den-piloten-*DAT-PL* | dem-kommandanten-*DAT-SG* | den-kommandanten-*DAT-PL*
16	V_{trans}	→	überrascht (0.05) | überraschen (0.05) | begeistert (0.2) | begeistern (0.2) | sieht (0.125) | sehen (0.125) | hören (0.125) | hört (0.125)
17	$V_{ditrans}$	→	zeigen | zeigt | empfehlen | empfiehlt
18	ADVP	→	MOD ADV | ""
19	MOD	→	äusserst | ""
20	ADV	→	überrascht | begeistert | freudig | glücklich
21	RP_{nom}	→	der-RP-*NOM-SG* | die-RP-*NOM-PL*
22	RP_{acc}	→	den-RP-*ACC-SG* | die-RP-*ACC-PL*
23	RP_{dat}	→	dem-RP-*DAT-SG* | denen-RP-*DAT-PL*

The grammar includes nominative, accusative and dative noun phrases, where each of the noun phrases takes a relative clause in 9% of the cases. Relative clauses (RC) are either subject-extracted (1/3 of RCs, SRC) or object-extracted (2/3 of RCs, ORC). Object-extracted relative clauses are either accusative (1/2 of ORCs) with either a transitive or ditransitive RC verb, or dative (1/2 of ORCs) with ditransitive RC verb. The grammar allows for (in principle unlimited) recursion of relative clauses of all three kinds. Verbs and nouns have to agree in number. To avoid unnecessary complexity due to ambiguity in gender and determiner-form, only masculine nouns are used, and are represented as determiner-noun clusters. Case and number are marked explicitly for nouns and relative pronouns to avoid the complex frequency biases of particular nouns due to differences in case marking between singular and plural. Sentences only include third person and present tense. Commas were not included. Table 4.2 shows examples of different sentences generated by the grammar.

Table 4.2: Examples of sentences generated by the PCFG

	type	example
1	MC, trans.	der-pilot-NOM-SG sieht den-kommandanten-ACC-SG EOS
2	MC, trans.amb.	der-pilot-NOM-SG überrascht den-kommandanten-ACC-SG EOS
3	MC, trans.amb + amb. ADV (GP)	der-pilot-NOM-SG überrascht begeistert den-kommandanten-ACC-SG EOS
4	MC, + amb. ADV	der-pilot-NOM-SG sieht glücklich den-kommandanten-ACC-SG EOS
5	MC, ditrans.	der-pilot-NOM-SG zeigt dem-kommandanten-DAT-SG den-piloten-ACC-SG EOS
6	SRC, trans.	die-piloten-NOM-PL die-RP-NOM-PL den-piloten-ACC-SG sehen hoeren die-piloten-ACC-PL EOS
7	ORCacc	der-pilot-NOM-SG den-RP-ACC-SG der-kommandant-NOM-SG sieht hoert den-piloten-ACC-SG EOS
8	ORCacc, + amb. ADV (GP)	der-pilot-NOM-SG den-RP-ACC-SG der-kommandant-NOM-SG überrascht sieht hoert den-piloten-ACC-SG EOS
9	ORCdat	der-pilot-NOM-SG dem-RP-DAT-SG der-kommandant-NOM-SG den-piloten-ACC-SG zeigt hoert den-piloten-ACC-SG EOS
10	ORCdat + amb. ADV (LSC)	der-pilot-NOM-SG dem-RP-DAT-SG der-kommandant-NOM-SG überrascht den-piloten-ACC-SG zeigt hoert den-piloten-ACC-SG EOS
11	double embedding	der-pilot-NOM-SG der-RP-NOM-SG den-kommandanten-ACC-SG denen-RP-DAT-PL die-piloten-NOM-PL den-kommandanten-ACC-SG zeigen hört den-piloten-ACC-SG EOS

Local syntactic coherence Locally syntactic coherent sequences are constituted by the two ambiguous elements «überrascht» and «begeistert», and occur in dative ORCs, (see table 4.2, line 6, for an example). To mirror the bias towards verb usage similar to the Finiteness Bias *FB* evaluated in experiment 2 (section 3.5), the probabilities in the grammar are adjusted in the following way: If a phrase structure includes a transitive verb on its right-hand side, one of the ambiguous verbs has a probability of .1 of being selected («begeistert»/«begeistern»; FB_{low}), while the other one («überrascht» / «überraschen» FB_{high}) is selected with a probability of .4. Unambiguous transitive verbs are selected in the remaining cases FB_{namb} (see table 4.1, line 16).

Thus, «begeistert» has a higher probability of occurring as a verb than «überrascht», compared to their overall frequency. Adverb use was equally distributed between all 4 ambiguous and unambiguous adverbs (see table 4.1, line 20). This results in a FB of .43 for «begeistert» and .16 for «überrascht». This particular way of implementing the difference in bias leads to a higher overall frequency of «begeistert» and «begeistern» ($freq_{beg} = 39436$), compared with «überrascht» and «überraschen» ($freq_{ueb} = 21796$). This frequency difference could be balanced by adjusting the selection probabilities at adverb positions. However, this would induce a systematical adverb bias that is not desirable.

Full-clause local syntactic coherence occurs in dative ORCs with an ambiguous adverb in the relative clause:

der-pilot-*NOM-SG* dem-*RP-DAT-SG* ...
the-pilot-*NOM-SG* whom-*RP-DAT-SG* ...

[LC:] der-kommandant-*NOM-SG* überrascht den-piloten-*ACC-SG*
[LC:] the-commander-*NOM-SG* surprisedly/surprises the-pilot-*ACC-SG*

... zeigt sieht den-kommandanten-*ACC-SG* EOS
... shows sees the-commander-*ACC-SG* EOS

The modifying adverb «äußerst» [very] (table 4.1, line 18 & 19) is included to model the design used in the visual world experiments. Thus, the factors *ambiguity of adverb (amb.adv)* and *additional adverb (add.adv)* can be used to evaluate local coherence effects similar to the visual world evaluation. The design will be described in more detail in section 4.3.1.

The hypothesis is that effects of local syntactic coherence can be understood as false alarm predictions based on local transitional probabilities that have been learned from frequent string sequences. These sequences will be called *source sequences*, as they provide the source for the incorrect predictions. The relevant source sequences are main clauses with (ambiguous) transitive verbs (see table 4.2, line 2, for an example).

The ambiguous elements do not only constitute local syntactic coherence in some sentences, but can also lead to temporal ambiguity and thus produce garden path sentences (see table 4.2, lines 3 & 8). This is always the case if the ambiguous element w occurs in a position n where both a transitive singular verb or an adverb is grammatical. In these cases, at position n it is not clear whether the element is used as a verb or as an adverb, which is only disambiguated at the next position $n+1$ (or, in some cases, even later): If w_{n+1} is a verb, w_n was used as an adverb; if w_{n+1} is a noun phrase, w_n was used as a verb.

Using the grammar described above, 5 corpora of 20000 sentences each were generated, using different random seeds. The relatively high number of sentences per epoch was chosen to make sure that each corpus contains a reasonable number of each of the different sentence types and variants, and to ensure that the distributional properties in the corpus approximate the probabilities specified in the grammar. Table 4.3 shows some of the distributional properties of the 5 training corpora.

Table 4.3: Properties of the 5 training corpora

	number words	sent.length mean	sent.length max	number RC	number amb.el high	number amb.el low	LC high	LC low	number source
co 1	127782	19	38	5180	5521	3722	43	38	1255
co 2	127669	18	32	5209	5445	3763	40	51	1279
co 3	128635	18	40	5372	5558	3744	41	42	1255
co 4	128758	19	35	5387	5537	3758	47	59	1216
co 5	128525	18	38	5378	5519	3791	54	61	1318
mean	128274	18	37	5305	5516	3756	45	50	1265
sum	641369			26526	27580	18778	225	251	6323
SD	509	0.48	3.1	102	43	25	5.7	10	37

number words: number of words in corpus; sent.length: sentence length
number RC: number of relative clause
number amb.el: number of ambiguous elements (high/low FB
LC high: number of sentences containing LSC, high bias;
LC low: number of sentences containing LSC, low bias;
number source: number of source sequences;

4.1.2 Architecture, Parameters and Procedure

Architecture and parameters were selected according to some "rules of thumb" that have been proven to yield reasonable results both in the literature and in a number of preliminary simulations (not reported here). Selection of the parameters were mainly based on the performance of the network in learning number agreement, since this task turned out to be most problematic to learn. A relatively large range of initial random weights has been proven to be a good guess for the task at hand (Rohde & Plaut, 1999, simulation 2). Since the grammar used is not too complex, moderate learning rate and moderate to high momentum was used. As momentum, I used *Doug's momentum*, which copes with the problem that a high momentum can lead to settling into local minima at the beginning of training.[1] In the following, I will only report simulations that were run with one set of parameters (table 4.4).

[1]http://tedlab.mit.edu/ dr/Lens/thumb.html: Weight Update Algorithms

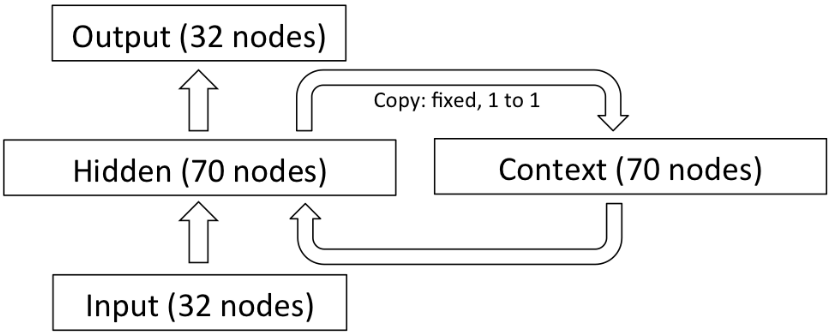


Figure 4.1: Architecture of the SRN

Architecture As architecture, I used simple recurrent networks with localist representation at the input and output layer. Each terminal of the grammar corresponds to one node in both input and output layer, resulting in 32 nodes in both layers. Hidden layer size was set to 70 units, roughly 2 times the number of input nodes, which has been proven as a reasonable size in several SRN models (MacDonald & Christiansen, 2002) and in our own preliminary simulations.

For each training corpus, 10 networks were initialized with different initial random weights. Each of the 50 networks was trained for 50 epochs, where *epoch* refers to the presentation of all 20000 sentences of the training corpus. Sentences were randomized before each epoch.

Table 4.4: Parameters of the SRN simulation

corpora	nets per corpus	runs	epochs	learning rate	momentum	initial range
5	10	50	50	.05	.8	.7

4.1.2.1 Training

The networks were trained with the *simple recurrent backprop through time* algorithm described in the LENS Manual[2], which yielded the best results regarding long-distance dependencies in the preliminary simulations.

Admittedly, backpropagation in general is not the strongest algorithm with regard to neural plausibility, and the particular variant was chosen for purely pragmatic reasons. It just yielded

[2] http://tedlab.mit.edu/~dr/Lens/elman.html#SRBPTT

better results than standard backpropagation. Similar to the parameters described above, the choice was exclusively based on performance on number agreement as a precondition to consider the networks as valuable models in the first place, and not on the effects of local syntactic coherence. Since the algorithm tends to give special status to the sentence as a training unit, it could be argued that this method is not the best one to be used to model the effects of local syntactic coherence, that can be said to reflect some degree of ignorance about this unit. However, as I have argued above, it is crucial to ensure that the networks are able to learn the grammar in the first place, and the chosen algorithm revealed the best results with regard to general performance. In addition, the likely weakening of local influences provides an even more conservative model with regard to my claims.

Output nodes employ a logistic activation function, and cross-entropy was used to calculate the output error that guides adjusting the connection weights. The task of the networks is to predict the next word in the sentence.

4.1.2.2 Testing

After each epoch, the networks were tested on different sets of test sentences. In the testing phase, the network was presented with the test sentences word by word, and the activation at the output layer was recorded for each input word. During testing, connection weights are fixed and no learning takes place.

To test general performance and success of learning, a random set of 5000 sentences was generated with the training grammar. I will refer to this test set as *test5000*. This particular test set will be used to gain an insight in to the general performance and the learning process of the networks, and, in particular, on performance on multiple embedded relative clauses.

A second test set was constructed manually and contains only sentences with one centrally embedded relative clause. I will refer to this test set as *testRC*. The test set was constructed in a way that all possible types of relative clauses (SRC, OCRacc, ORCdat) are included in several variants. This involved combining and tagging the following variables:

- *type*: SRC, ORCacc, ORCdat
- *adv*: RC contains adverb (yes/no)
- *bias*: VFbias of adverb (high/low/unambiguous)
- *ambig*: adverb is ambiguous (yes/no)

- *add.adv*: RC contains an additional modifying adverb (yes/no)
- *rc.ditrans*: RC verb is ditransitive (yes/no)
- *rcv.ambig*: RC verb is ambiguous (yes/no)
- *mcv.ambig*: matrix verb is ambiguous (yes/no)
- *num.mc*: number of matrix verb (SG/PL)
- *num.np2*: number of second NP (SG/PL)
- *num.np3*: number of third NP (SG/PL)

The complete test set *testRC* contains 896 sentences (SRC: 336; ORCacc: 336; ORCdat: 224) Subsets of *testRC* will be used to carry out further evaluation of performance on number agreement (section 4.2) and two experiments on LSC effects (section 4.3.2).

The particular test sets and the design for the experiments will be described in detail in the corresponding sections.

4.2 Evaluating the Plausibility of the Model: Long Distance Dependencies

In the following section I will evaluate whether the networks learn the crucial properties of the grammar at all, which is a necessary condition to consider the model as a plausible *cognitive* model in the first place. If not, any effects elicited by the model, whether or not they match with behavioral results, are questionable and could be mere artifacts of insufficient learning success.

The property I will use to evaluate learning success is the accuracy in long-distance number agreement. This task is particularly suited to evaluating the validity of the model for two reasons.

1. Number agreement of distant elements, potentially separated by multiple embedded relative clauses, is arguably the most complex task in acquiring the grammar. However, the processing of long-distance dependencies is a well-known weakness of standard SRNs as they are used here (cf. Hochreiter & Schmidhuber, 1997), but constitutes a crucial

property of human languages. Thus, it constitutes a good testbed for the power and limits of the model.

2. With regard to local syntactic coherence, capturing long-distance dependencies is particularly important for another reason. If we want to evaluate interference of local and global context, we have to ensure that the networks do not operate *solely* on a limited local context – otherwise, effects of local coherence would be a mere artifact of the fact that the networks are not able to take long-ranging regularities into account at all.

Since the training grammar is recursive, in principle it allows for an unlimited number of embedded relative clauses between the subject of the matrix clause and the matrix verb. With regard to the task the SRNs perform, processing long-distance number agreement comes down to activating only verbs that match in number with the respective noun. The distance between the dependent elements increases considerably with the level of embedding, rendering accurate prediction a particularly hard task to learn for the networks.

In the following section, I will show that the networks are able to perform the task to a reasonable degree, although not perfectly. However, difficulty or even failure to process sentences with multiple levels of embedding is a well-known phenomenon in human behavior (e.g. Gibson & Thomas, 1999). In addition, in natural language the "in principle" unlimited amount of recursion is strongly limited (Karlsson, 2007). Thus, I will argue that the networks can be taken as a reasonable partial model of human language capacities both with regard to their power and their limits.

4.2.1 Methods of Evaluation

SRNs are trained on predicting the next word in a sentence. Given the localist representation of input and output, the training vector $\mathbf{T}_i$ is a bit vector where the node representing the actual next word is set to 1 and all other nodes to 0. At the majority of sentence positions, however, there are several grammatically legal continuations, for example the different members of a word class, and thus the network will activate all of them to some degree. Therefore, to compare the value of output activations $\mathbf{A}_i$ with the training vector $\mathbf{T}_i$ which represents only the actual next word, is not very useful to evaluate learning success, because the error will stay relatively high even if the network predicts only legal continuations.[3] Learning the grammar

[3]Comparing the training vector and the output values of the actual next word, however, can be used to calculate SRN based surprisal values and thus derive a measure of, for example, reading time (Frank, 2009).

comes down to activating any lexical item to the degree that this item is to be expected, according to the training grammar. In other words, the output activation vector $\mathbf{A}_i$ should reflect the transitional probabilities of each element at the current position. This vector of transitional probabilities can be derived from the PCFG underlying the training materials, and is provided by the tool *SLG* that was used to generate the materials. The vector will be referred to as *grammatical probability vector* $\mathbf{G}_i$.

Several measures can be used to calculate the error by comparing output activation vector $\mathbf{A}$ and grammatical probability vector $\mathbf{G}$. Distance measures such as *Root mean squared error* or the *cosine*, representing the angle between vectors, are commonly used to grasp the deviation between output activation and target values and thus evaluate network performance.

A measure that was explicitly developed to evaluate SRN performance in the prediction task is the *grammatical prediction error GPE* (Christiansen & Chater, 1999b). The *GPE* takes into account *hits* (correctly activated nodes), *misses* (no or to little activation of correct nodes) and *false alarms* (activation of incorrect nodes or too high activation of correct nodes). It takes a value of zero if all legal elements are activated according to the grammar and all non-legal elements are not activated (see MacDonald & Christiansen, 2002, p. 54, for a detailed description). The *GPE* and similar error measures provide useful information about the time course of learning and can be used to compare performance of different architectures and parameters. They are also insightful in evaluating the general difficulty SRNs exhibit on different sentences types or sentence positions. The disadvantage of any distance measure of this kind, however, is that the error value does not tell us much about what exactly produces the error (Ruh et al., 2002; Konieczny, Ruh, & Müller, 2009). Similar error scores in different situations can result from various sources that can not be easily distinguished. In addition, intermediate values (the *GPE* ranges from 0 to 1) are not readily interpretable as "good" or "bad", but can only provide information about performance in relation to other contexts, architectures or parameter settings.

4.2.1.1 Comparison of SRNs With Other Language Models

A possibility to overcome this problem and to evaluate what a network is capable of is to compare a network's error score to the error score of another "device" that is known to have particular abilities and lack others. The comparison of SRN output with PCFG-based transitional probabilities described above can be viewed as a comparison between different language models. Since in our case the PCFG provides the correct language model, perfect match (zero error) would thus signify that the SRN implements the underlying grammar. However, if we want to know which aspects of the underlying grammar are not learned by the SRN, it is more useful to compare its performance to different language models.

This approach has been used in several studies, where SRN performance has been compared to the performance of n-gram models (e.g., Christiansen & Chater, 1999b; Cho, Szkudlarek, Kukona, & Tabor, 2011). In this work, I will compare SRN performance to two different kinds of models, n-gram models of various length and a language model that implements all properties of the grammar except number agreement. This second model will be called the *number ignorant* model *NI*.

N-gram models N-gram models of various length are useful benchmarks to prove if the networks merely act on a "shortsighted" window of sentential context, or if they take into account general properties of the underlying grammar as the recursive embedding of subordinate clauses. In many cases, a narrow context of several words will produce good results. With regard to number agreement, however, n-gram models that take into account only a context of n words, will obviously perform very poorly as soon as the distance between the dependent elements is larger than n. Thus, if the networks perform better than an n-gram model at a particular position, we can be confident that the SRNs take into account a context larger than n words. In addition, n-gram models are also useful to detect and visualize biases in the training materials that are not easily inferred from inspecting the probabilistic grammar (see, for example, section figure 4.11).

To calculate n-gram based transitional probabilities, a corpus of 50,000,000 sentences was generated with the training grammar (table 4.1). Within this corpus, word sequences of length $n = 1$ to $n = 5$ were counted, and n-gram based transitional probabilities for the test sentences were calculated according to equation (4.1). This rather large number of sentences is necessary to generate enough n-gram sequences to represent the distributional properties resulting from the grammar.[4]

$$TP^n(w_i) = P(w_i|w_{i-n}\dots w_{i-1}) = \frac{freq(w_{i-n}\dots w_i)}{freq(w_{i-n}\dots w_{i-1})} \tag{4.1}$$

[4]It should be noted that the n-gram counts do not represent the distributional properties of the particular materials any of the networks have been trained on, but the distributional properties resulting from the grammar (at least, due to the large number of sentences, a rather close approximation of the properties of the grammar).

For any position i in a sentence, $TP^n(w_i)$ is calculated for each lexical item w, resulting in a vector $\mathbf{N}_i^n$ representing the probability distribution over lexical items predicted by the $n-1$ preceding words at position i. $\mathbf{N}_i^n$ can then be compared to the grammatical probability vector $\mathbf{G}_i$ by calculating the grammatical prediction error GPE_i^n of the n-gram model at position i. If the network's error GPE_i^{SRN} is below GPE_i^n, it signifies that the network takes into account a broader context than n words.

Number ignorant model *NI* Comparison with n-gram models can provide fruitful insights into the power and limits of networks' abilities. However, to evaluate performance on a particular task, such as number agreement, it is helpful to construct a model that only fails with respect to this particular task, while doing everything else correctly. With regard to number agreement, such a model will predict the correct word classes at each point in the sentence with accurate probabilities according to the grammar. However, at positions where number agreement has to be respected, it will "guess" the number of verbs, thus assigning equal values to both singular and plural forms of (legal) verbs. The transitional probability TP_i^{NI} of a word w at position i as predicted by *NI* can be derived from $\mathbf{G}_i$ and provide a probability vector $\mathbf{NI}_i$ over all lexical entries. For a simple grammar, $\mathbf{NI}_i$ can be derived by taking the probability vector $\mathbf{G}_i$ and, for each grammatical verb, take its probability and assign half of this value to each singular and plural form of the verb.

The presence of ambiguous elements in the grammar, however, requires some additional effort. Consider, for example, the word «überrascht» which can occur either as an adverbial element or as a transitive singular verb. As a result, the transitional probability of the element at position i is determined by both probability of adverbs and probability of (transitive singular) verbs. Since the number ignorant model *NI* predicts adverbs correctly, the portions of probability due to adverbial and verbal use have to be separated. In the grammar, all adverbs are equally probable at a given position (see table 4.1, line 20), therefore the probability value of one of the unambiguous adverbs TP_{ADV}^G can be used to determine the adverb portion. Subtracting TP_{adverb}^G from $TP_{ueberrascht}^G$ then gives us the verb portion, which can then be divided by two and assigned to both forms of the verb, «überrascht» and «überraschen». The adverb portion is then added to the resulting value of the ambiguous form «überrascht». For unambiguous verbs, the probability of the correct verb form is divided by two and assigned to singular and plural forms. Given the (hypothetical) case that at position i «überrascht» has a probability $TP_{ueberrascht}^G$ of 0.5 and an unambiguous adverb has a probability TP_{ADV}^G of .1, $TP_{ueberrascht}^{NI}$ will be .3, while $TP_{ueberraschen}^{NI}$ will be .2.

Transitional probabilities will also differ for transitive and ditransitive verbs in some sentence positions, since, for example, the RC verb of a dative ORC has to be ditransitive. Finiteness bias *FB* of ambiguous elements also influences transitional probabilities of verbs at particular sentence positions. These differences are accounted for by calculating TP_v^{NI} for each particular verb v separately. Values of all other terminal nodes are set to their transitional probabilities according to $\mathbf{G}_i$. Thus, for each position i in a sentence, we can generate a vector $\mathbf{NI}_i$, representing predictions of *NI* at position i.

Calculating GPE_i^{NI} from $\mathbf{NI}_i$ and $\mathbf{G}_i$ will tell us what the error would be if a network has learned the correct categories, but not number agreement. It can thus be used as a baseline that networks' errors can be compared against. Notice, however, that this measure is rather conservative, since all activations except the verbs' perfectly match the grammatical probabilities. In contrast, the networks will never reach a perfect match for any element. As a consequence, if the GPE^{SRN} is below GPE^{NI}, it can be taken for granted that number agreement has been learned.

However, a further problem has to be taken into account. In preliminary simulations I encountered the problem that averaging error scores over a set of different test sentences can lead to rather overall error scores if one type of sentence has been learned fairly well, while another type has not been learned at all. Networks seemed to perform fairly well in processing long-distance number agreement, according to the error scores. However, detailed analysis showed that the networks were almost perfect only in singular sentences while only at chance level in plural sentences. As the result, since the network was almost perfect in 75 % of the cases (all singular and half of the plural sentences), and in the remaining 25 % only predicted verbs and no other elements, the averaged *GPE* was very low overall – but the network obviously learned something very different from number agreement. Thus, averaged error scores do not necessarily reveal if number agreement has been learned. Instead, it is important to evaluate if number agreement is achieved or not *for each particular sentence*. In the next section, I will describe a method that allows us to do this by taking into account both transitional probabilities **G**, output activations **A** and the predictions of the number ignorant model **NI**.

4.2.1.2 How to Evaluate Number Agreement Performance?

The underlying idea of the measure is that a network can be said to achieve number agreement if and only if the pattern of output activations of the singular and plural form $\mathbf{A}_v$ of each verb pair (singular and plural form) mirrors the corresponding pattern of transitional probabilities $\mathbf{G}_v$.

The crucial property thus is the relation between the pattern of transitional probabilities of the singular and plural forms of a verb, and the corresponding pattern of output activations. The patterns for each verb pair can be represented as the slope of a line connecting singular and plural value of a verb. For each verb pair v, the slope s is calculated for output activations (s_v^A), the PCFG-derived transitional probabilities (s_v^G), and the predictions from the number ignorant model (s_v^{NI}).

If s_v^A is between s_v^{NI} and s_v^G, the network activates the singular and plural nodes of the verb in the correct pattern, and the verb pair can be counted as correctly activated. Taking into account only the slopes assures that a pattern can be counted as correct or incorrect even if the actual activation values differ from the transitional probabilities, for example, due to other properties than number agreement that may have not been learned perfectly.

Figure 4.2 demonstrates this method with made-up values. Output activations $\mathbf{A}_v$ are plotted as a blue line, transitional probabilities $\mathbf{G}_v$ as a green line, and the predictions of the number ignorant model $\mathbf{NI}_v$ as a red line. The upper panels show values, while the lower panels only represent the slopes. In the left-hand column, the slope of activation s_v^A and probabilities s_v^G are both positive, while the slope of the number ignorant prediction s_v^{NI} is 0. Thus, it is an example of a verb where singular and plural forms are activated in the correct pattern and is thus considered as a *hit*. This can be seen more clearly in the lower panel, representing only the slopes: The blue s_v^A line is in between the green s_v^G and red s_v^{NI} line.

The right-hand column represents an example of failed agreement: The blue line's slope s_v^A is smaller than the red line's slope s_v^{NI}, while the green line s_v^G has a positive slope, indicating that the verbs are activated incorrectly. Note that the red s_v^{NI} line does not have a slope of 0 in this case, which happens for ambiguous singular forms, where the value represents both adverb and verb probability. This exemplifies that it is not sufficient to just compare the direction of $\mathbf{A}_v$ and $\mathbf{G}_v$ slopes, but that the "number ignorant" prediction $\mathbf{NI}_v$ has to be taken into account as a baseline.

The comparison of slopes results in a binary (hit/miss) measure for each verb pair. For a particular sentence, number agreement is conceived as correct if all verb pairs are counted as hits.

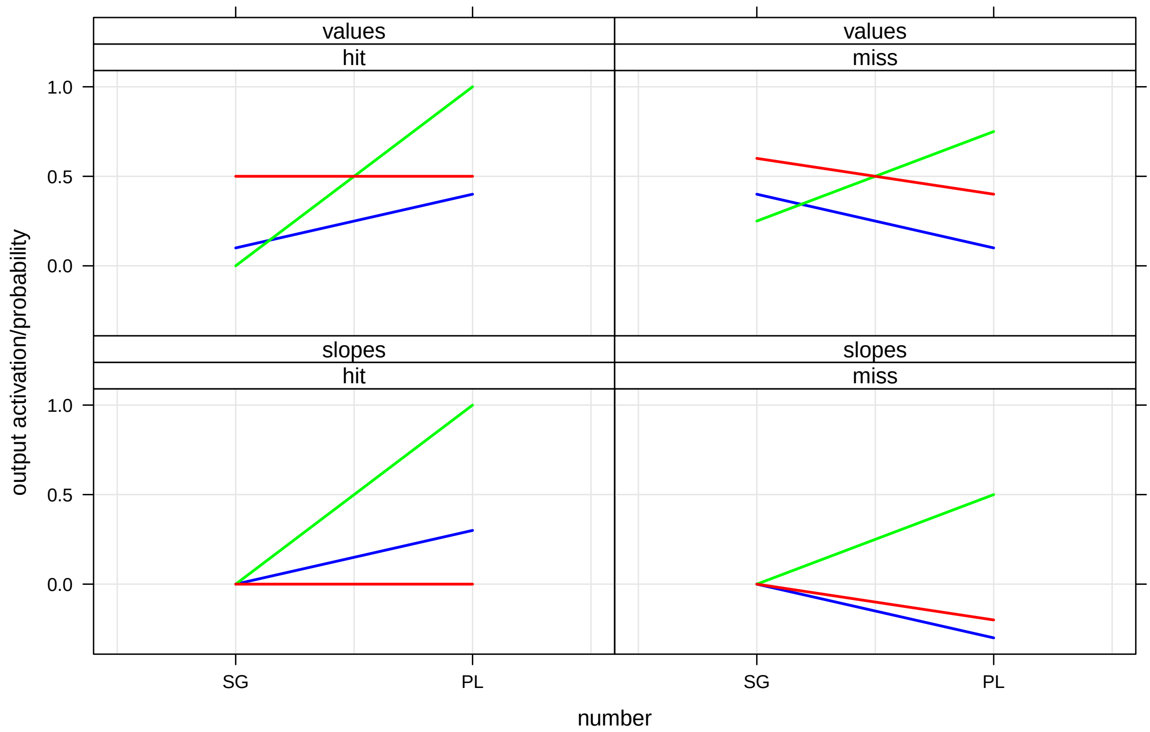


Figure 4.2: Visualization of the measure used to evaluate number agreement (data is made up for explanatory purposes). Red lines represent prediction of the number-ignorant device, green lines transitional probability derived from the PCFG (target), and blue lines output activations. In the left column, number agreement is correct: the slope of the blue line is in between slopes of green and red line. The right column represent a miss, since the blue line's slope points in the wrong direction. The upper panel includes activation/probability values, while the lower panel only represents the slopes.

4.2.2 Results

The results section will proceed as follows: I will first inspect two individual sentences of different complexity to give an impression of how the networks perform and how sentence complexity influences performance. I will then evaluate performance in number agreement in the larger test set with 5000 sentences (*test5000*), focusing on depth of embedding as a determinant of complexity. Since I will use single embedded relative clauses to investigate effects of local coherence later on, I will conclude the section with a closer look at sentences within this class.

4.2.2.1 Two Examples

To give an impression of the networks' learning process, capabilities and limits, I choose two sentences from *test5000*. First, we will have a look at a rather simple sentence with one centrally embedded SRC:

(46) MC der-pilot-*NOM-SG*
SRC der-*RP-NOM-SG* die-piloten-*AKK-PL* ausserst begeistert
RCV begeistert
MV ueberrascht
MC die-kommandanten-*AKK-PL* eos

Figure 4.3 shows the *GPE* scores on each word in the sentence, and their development over training. GPE^{SRN} scores are averaged over all 50 networks (thus treating individual networks as if they were participants in an experiment). The colored lines represent states of training, ranging from epoch 1 (dark blue) to epoch 50 (red). GPE^{SRN} of the untrained network is plotted in black as a baseline. N-gram based errors GPE^{n} ($n = 3$ to $n = 5$) are plotted as step lines. While at most positions GPE^{SRN} is rather low (between .2 and .4), the error is significantly higher at both the RC verb and the Matrix verb, signifying that the verbs are the most difficult positions to learn, since not only the correct category has to be predicted, but also the correct number of the verbs. With training, however, error at the verb positions drops and reaches the level of the other positions. The difficulty of the verb positions is also mirrored by the high scores of GPE^{n}, in particular at the matrix verb. At the matrix verb, GPE^{SRN} drops below 5-gram error already after only one epoch of training, indicating that the network is not restricted to a shortsighted window of words. We can also see that GPE^{SRN}, in contrast to GPE^{n}, never reaches zero, showing that the networks never achieve a perfect

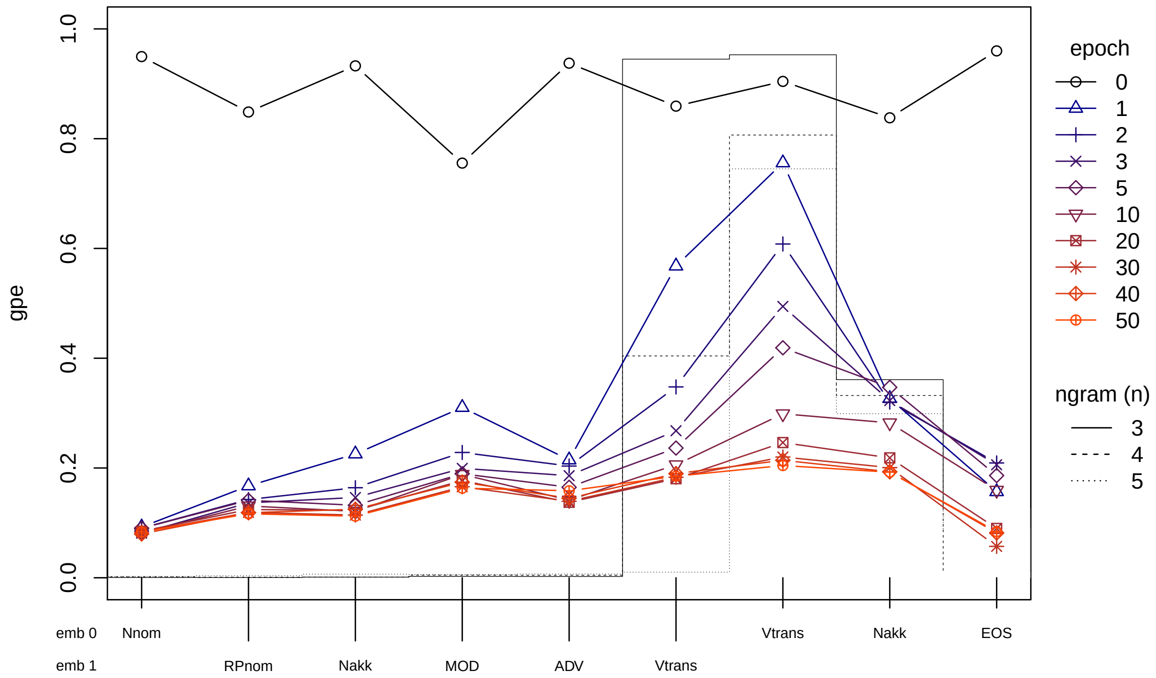


Figure 4.3: Example of Grammatical prediction error in example sentence with 1 embedded SRC. Black line shows GPE^{SRN} before training. Steplines represent ngram-based prediction error GPE^{n}.

match to the grammatical probabilities. This is noteworthy insofar as it shows that comparing error values of networks with "partially perfect" models like n-gram models is often rather conservative. For these purposes here, however, the comparison is sufficient to show that the networks outperform n-gram models and grammar learning has been achieved, at least for this sentence.

The picture, however, is very different if we take a look at the following sentence:

(47) MC der-pilot-*NOM-SG*
ORC 1 dem-*RP-DAT-SG* der-kommandant-*NOM-SG* ueberrascht die-kommandanten-*AKK-PL*
ORC 2 die-*RP-AKK-PL* die-kommandanten-*NOM-PL*
ORC 3 die-*RP-AKK-PL* der-kommandant-*NOM-SG* freudig dem-piloten-*DAT-SG*
RCV 3 zeigt
RCV 2 hoeren
RCV 1 zeigt
MV sieht
MC äusserst begeistert den-kommandanten-*AKK-SG* eos

This sentence contains three centrally embedded relative clauses: a dative RC modifying the subject, an accusative RC modyfying the object of the dative RC, and a third accusative RC modifying the object of the second accusative RC.

Not surprisingly, the networks have considerable difficulty in processing this sentence: 14 words are in between the verb of the matrix clause «sieht» and the dependent subject «der-pilot-*NOM-SG*». Figure 4.4 shows the GPE^{SRN} at each position as well as the ngram-based prediction error GPE^{n}.

While network error at least decreases over training for the verbs of the embedded RCs, at the matrix verb, no clear effect of training can be observed. The error scores in later stages do not drop below the error produced by the untrained network. It is also observable that error increases with depth of embedding at the verb positions and is at least lower than 5-gram based error at some of the verbs. In conclusion, this sentence seems to be beyond the power of the networks.

However, although the comparison with n-gram based errors reveals some information about power and limits of the model, it does not really tell us much about what the network actually does and does not.

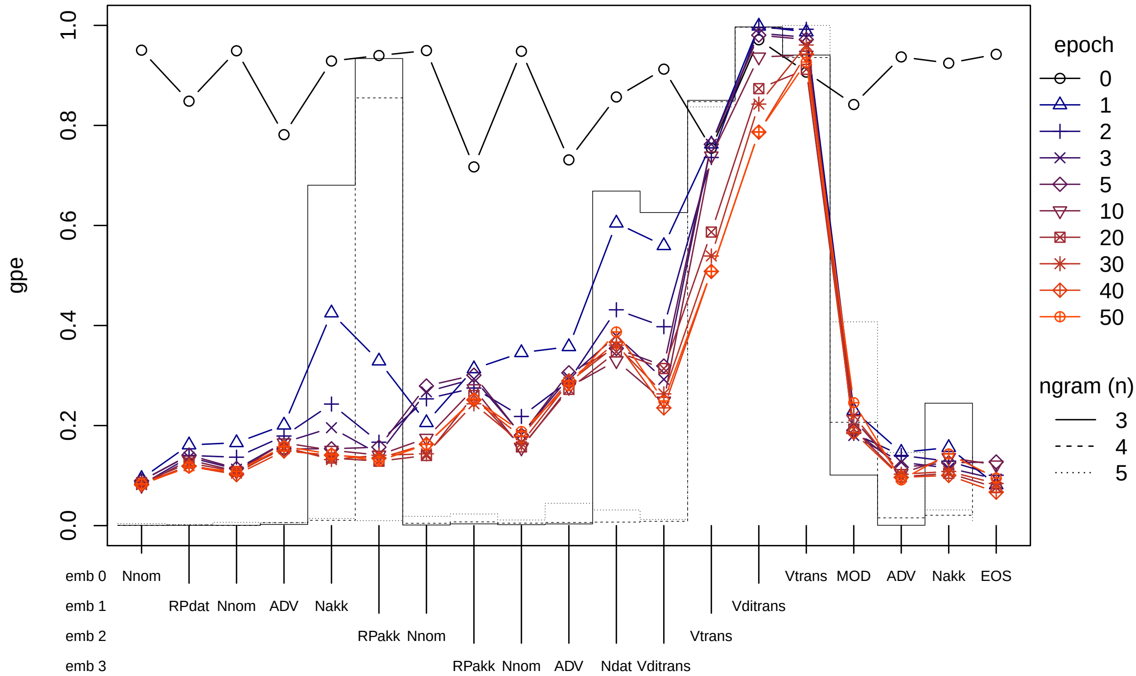


Figure 4.4: Example of Grammatical prediction error in example sentence with 3 levels of embedding. *GPE* values at verb positions increase with level of embedding. GPE^{SRN} at the matrix verb did not decrease after 50 epochs of training. Black line shows GPE^{SRN} before training. Steplines represent ngram-based prediction error GPE^{n}.

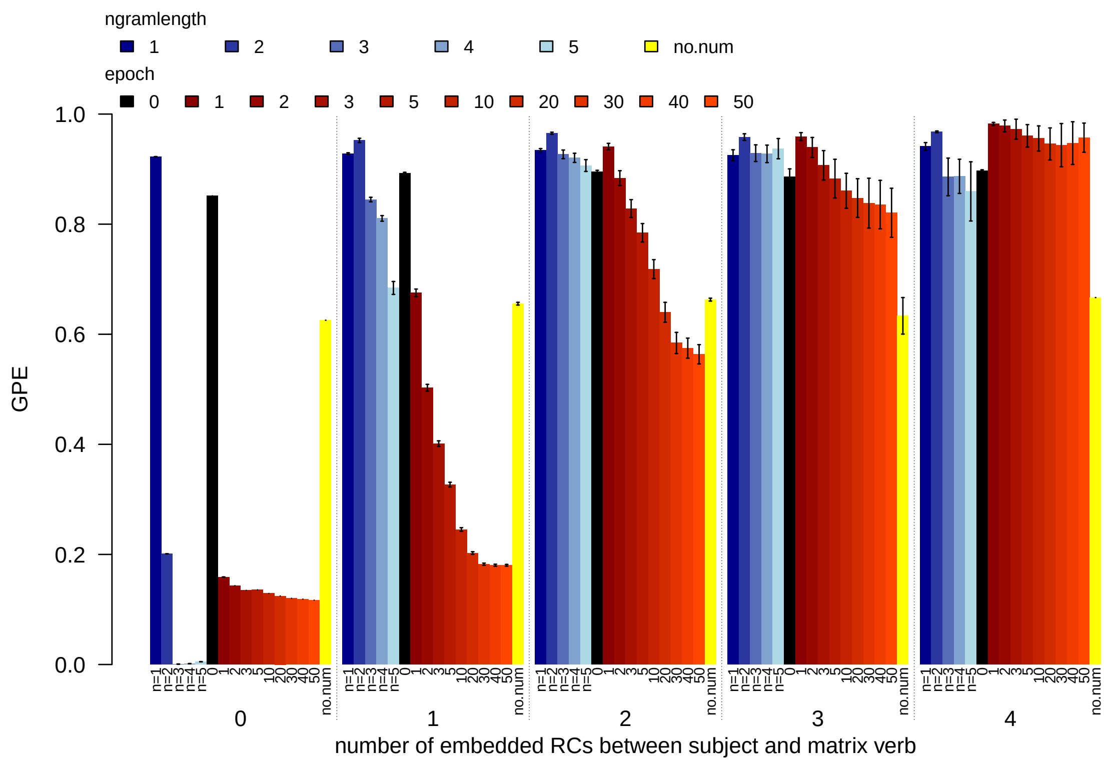


Figure 4.5: Grammatical prediction error, averaged over all (50) runs, at matrix verbs in sentences with different (0-4)levels of central embedding. Blue bars show ngram-based prediction errors. Red bars show networks' prediction error. Black bars show error before training (epoch 0). The yellow bar (no.num) shows the error of the number ignorant model *NI*. Whiskers indicate standard errors, due to different sentences. The larger standard error in deeper embeddings reflects that the training set (*test5000*) contains fewer sentences with deep embedding.

4.2.2.2 Influence of Depth of Embedding

I will now evaluate whether the pattern observed in the two individual sentences is reflected in the training set. Since the region of the matrix verb appears to be most interesting with regard to performance and learning success, I will focus on this position in the remainder of this section. Sentences in *test5000* are grouped by depth of embedding, counting the number of centrally embedded relative clauses between subject and matrix verb (*MC_rcdist*). A value of 0 indicates absence of a relative clause between subject and matrix verb. Example (46) would thus have a *MC_rcdist* value of 1, example (47) a MC_rcdist value of 3.

Figure 4.5 shows network GPE^{SRN} at the matrix verb in different training stages as red bars, separated by the number of relative clauses between subject and matrix verb. As comparison, n-gram based ($n = 3$ to $n = 5$) prediction errors GPE^n are plotted as blue bars, and the error produced by the number ignorant model GPE^{NI} as a yellow bar. The black bar represents the *GPE* of the untrained networks. GPE^{SRN} drops quickly in simple sentences (0) and for level 1 sentences, where networks clearly outperform both n-gram models and the number ignorant model. For level two sentences, it takes about 20 epochs of training until networks show lower error values than the number ignorant model. For level three sentences, although the networks get better than the n-gram models, they do not reach the level of the number ignorant model even after 50 epochs. Finally, in level 4 sentences the networks don't make any progress at all, producing errors that are even higher than the error of the untrained networks. The GPE^{SRN} value of almost 1 indicates that the networks, bluntly stated, have no clue what to predict.

In conclusion, networks' performance shows a clear effect of embedding depth, where performance is good in simple sentences and level 1 sentences, and at least reasonable in level 2 sentences. Level 3 and, more clearly, level 4 sentences seem to be beyond the capabilities of the networks.

As mentioned above, the comparison of global error values is useful to assess learning development, but does not tell us exactly what causes the error, although the comparison with the number ignorant device at least gives some indication. I will now have a look at number agreement more directly by using the slope comparison method described above. For each sentence in *test5000* and all 50 networks, output activation, PCFG probability and *NI*-prediction for each verb-pair are compared at the matrix verb position. Number agreement in a sentence is considered as correct in all cases where the activation slope points in the correct direction for all verb-pairs.

Figure 4.6 shows the proportion of hits, averaged over individual networks for each level of embedding. Each grey line represents the averaged proportion of hits for one sentence in the test set. A smoother (blue line) is added to represent the general tendency over sentences. As we can see, for level 1 sentences, the majority of sentences is processed correctly, although some sentences perform considerably worse. Proportion of hits in level two sentences is clearly worse, with a group of sentences that are virtually never correct. In level 3 sentences, none of the grey lines reach a value of 1, even after 50 epochs of training, indicating that none of the level 3 sentences are processed correctly by all of the 50 networks. Not surprisingly, level 4 sentences hardly show any progress during training, where the consistent value of 0 of one of the sentences indicates that none of the networks were able to process it correctly. Note that this measure is rather conservative, since also a case where 5 of 6 verbs are activated correctly is considered as a miss.

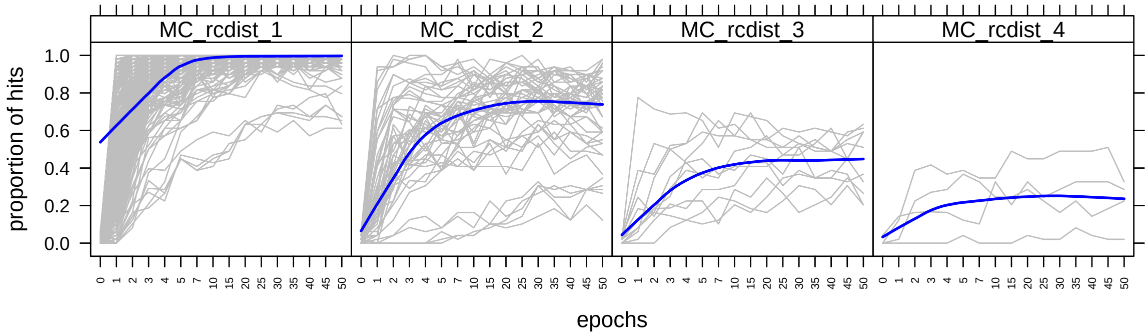


Figure 4.6: Proportion of hits (number) at matrix verb, averaged over networks. Testset consists of 5000 randomly selected sentences. Each grey line represents one *sentence*. The blue line is a smoothed fit (span=1/2)

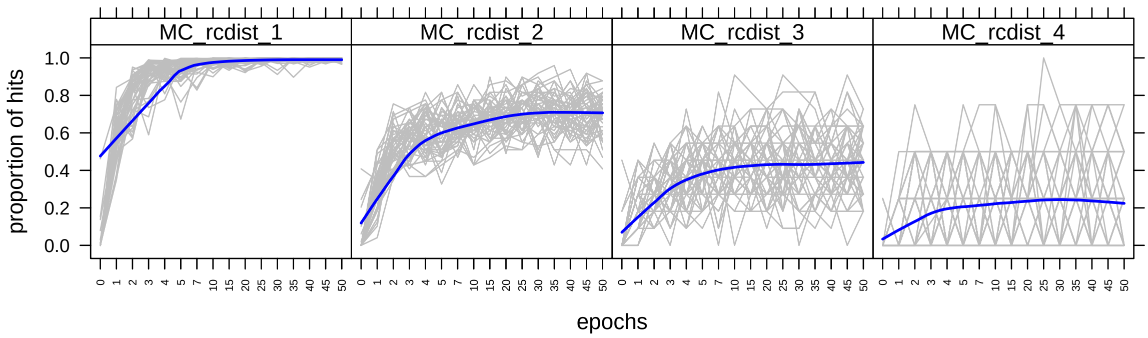


Figure 4.7: Proportion of hits (number) at matrix verb, averaged over sentences. Testset consists of 5000 randomly selected sentences. Each grey line represents one *network*. The blue line is a smoothed fit (span=1/2)

Averaging over sentences instead of networks, figure 4.7 shows that the variability between individual networks grows with depth of embedding. A value of 1 indicates that a network processed all sentences in the group correctly, which is the case for most networks in level 1 sentences. For level 2 sentences, no network gets all sentences correct, and variability between networks is considerably higher. The same holds true for level 3 and 4 sentences, where a considerable amount of data points with a value of zero indicates that some networks did not get any of the sentences correct.

To conclude, processing of long-distance dependencies is dramatically influenced by the distance between dependent elements, with a substantial breakdown if more than two relative clauses are embedded between them. The results clearly demonstrate that our model is not able to fully grasp the recursive nature of the grammar. On the other hand, the results also showed that the networks outperform n-gram models of at least $n = 5$, indicating that the SRNs do not operate exclusively on a shortsighted window of several elements. Given the complexity of the sentences and considering the evidence that humans perform considerably worse or even lose track in sentences with multiple embeddings, the model can be taken as a plausible candidate for a partial model of human sentence processing.

4.2.2.3 Different Types of Relative Clauses

To evaluate the effects of local syntactic coherence, we will use only sentences with a single embedded relative clauses, resembling the stimuli used in visual world experiment 2 (section 3.4). I will now have a closer look at this group of sentences to make sure that potential effects of local syntactic coherence are not mere artifacts of a fundamental incapability of the networks to process the sentences at all. To do this, I will use the manually constructed test set *testRC*, which only includes a variety of sentences with one embedded relative clause (see section 4.1.2.2).

Figure 4.8 shows the proportion of hits according to the slope comparison method at the matrix verb for single embedded sentences, grouped by type of relative clause. Data is averaged over networks. We can see that in general, ORCs take more training to be learned than SRCs. In addition, while in SRCs and dative ORCs the proportion of hits is close to 1 after about 20 epochs of training, some particular sentences with accusative ORCs are not processed correctly even after 50 epochs of training. A closer look reveals that these sentences contain ambiguous elements as the RC verb, which leads to a temporal ambiguity and a particularly complex pattern of transitional probabilities at the matrix verb. One particular sentence shows

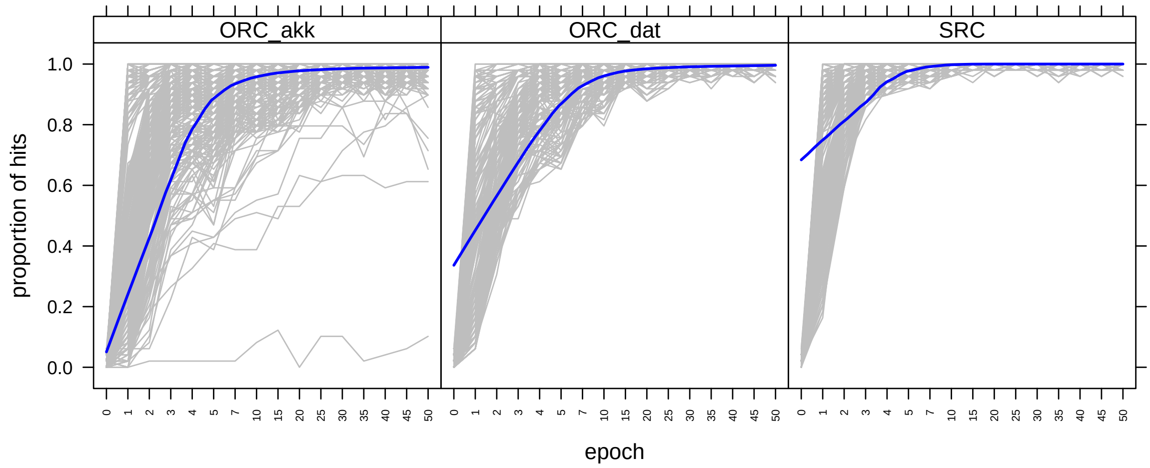


Figure 4.8: Proportion of hits (number) at matrix verb, averaged over networks. Each grey line represents a sentence, The blue line is a smoothed fit (span=1/2)

particularly bad performance, so we will have a closer look at this item, and contrast it with a randomly selected sentence where number agreement has been achieved. As an example of achieved number agreement, we take the SRC,

(48) der-pilot-*NOM-SG* der-*RP-NOM-SG* die-kommandanten-*AKK-PL* hoert begeistert den-piloten-*AKK-SG* EOS

Figure 4.9 shows transitional probabilities (green line) and the prediction of the number-ignorant device (red line). The three blue lines represent the output activations of three different networks (trained on the same corpus, but starting with different initial random weights). Rows represent different training levels, ranging from 0 (before training) to end of training after 50 epochs. Each column represents one of the 6 verb pairs in our grammar. Rows represent different training levels, ranging from 0 (before training) to end of training after 50 epochs. The slopes of the blue lines, starting at random values before training, approximate the slope of the transitional probabilities and thus change in the correct direction. In this example, the output pattern for all verb-pairs is correct already after one epoch of training, leading to a match value of 1.

The sentence showing worst performance in the whole test set is the following.

(49) die-piloten-*NOM-PL* die-*RP-AKK-PL* der-kommandant-*NOM-SG* überrascht überraschen den-piloten-*AKK-SG* EOS

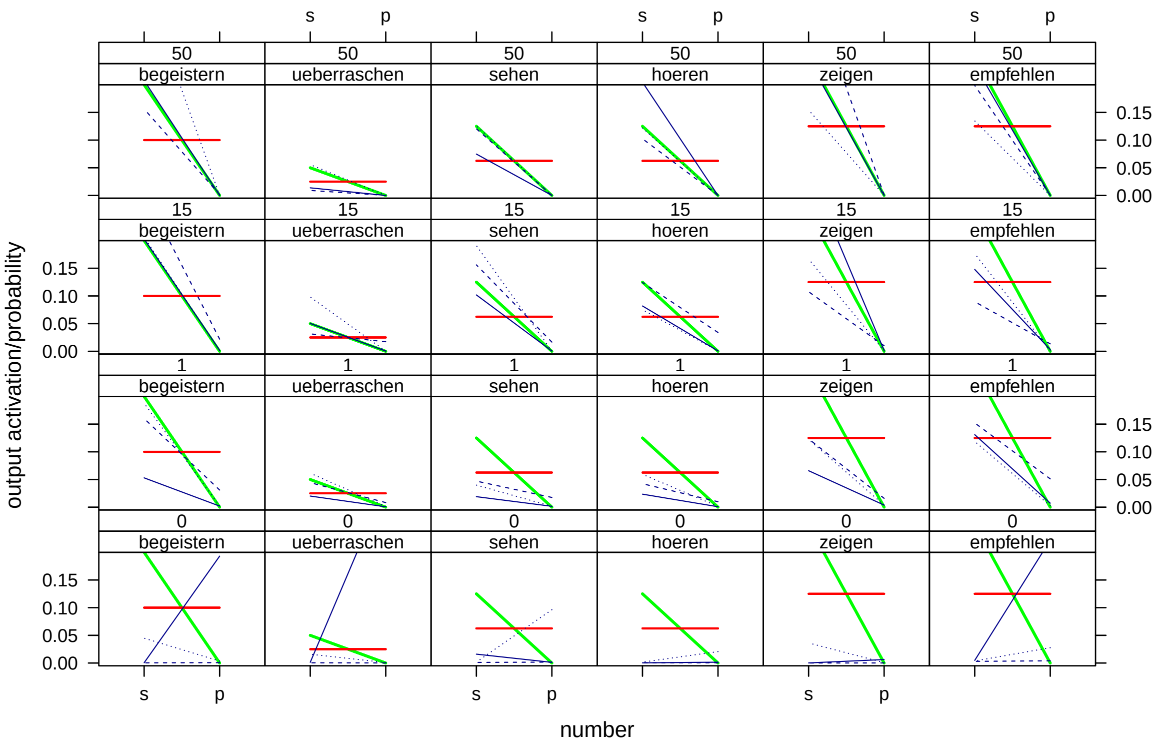


Figure 4.9: Output activation, transitional probability and NI-predictions of verbs at matrix verb position. Output activations of three different networks are plotted. Rows demarcate singular/plural pairs of each of the verbs used in the grammar. In this example, all networks activate all the elements in the correct pattern (slope), and will thus receive a match value of 1.

On first view, the sentence does not seem to be extraordinarily complicated, since neither an adverbial phrase nor a dative object is contained in the relative clause, which would be expected to make number agreement harder since the distance between dependent elements is prolonged. However, figure 4.10 reveals that the difficulties arise from ambiguity of the RC verb and the resulting pattern of transitional probabilities. As we can see, the activation slopes differ wildly between both verb pairs and networks, even after 50 epochs of training. Crucially, the patterns of transitional probabilities (green line) have different slopes for transitive and ditransitive verbs: The ambiguous element «überrascht» at the RC-verb position could be an adverb, as in

(50) die-piloten-*NOM-PL* die-*RP-ACC-PL* der-kommandant-*NOM-SG* ueberrascht → sieht ...
MC-subject-PL RC-subject-SG **ADV** → RC-verb-**SG**

Thus, transitive singular verbs are grammatical. However, «überrascht» can be the RC verb as well, in which case it would be followed by the (plural) matrix verb.

(51) die-piloten-*NOM-PL* die-*RP-ACC-PL* der-kommandant-*NOM-SG* ueberrascht → sehen ...
MC-subject-PL RC-subject-SG **RC-verb-SG** → MC-verb-**PL**

Thus, for transitive verbs, both singular and plural verbs are legitimate to some degree, and probabilities are even higher for singulars due to the finiteness bias *FB* of «überrascht». Ditransitive verbs (columns 5 & 6 in figure 4.10), in contrast, are legal continuations only as matrix verbs, since no indirect object occurs in the RC. Therefore, ditransitive verbs are legitimate only in plural. Thus, the failure of the networks to achieve number agreement in this particular sentence can be ascribed to the temporal ambiguity induced by the ambiguous RC verb. In particular, the fact that the probabilities of singular and plural differ for different verb classes not only in strength, but also in direction, renders this sentence particularly hard. The sentence is an interesting example of the complex influence of ambiguous elements on transitional probabilities.

4.2.3 Preliminary Discussion

The results reveal that the networks are capable of predicting the correct number of verbs in simple sentences, as well as in the majority of sentences with one embedded relative clause after rather short training. Level two sentences reach at least a reasonable degree of correctness, while level 3 and 4 seem to be mainly beyond the limits of network capability. For level 3, the networks still perform better than 5-gram models, and, after about 20-30 epochs of training, show smaller errors than the number ignorant device. However, using the slope comparison

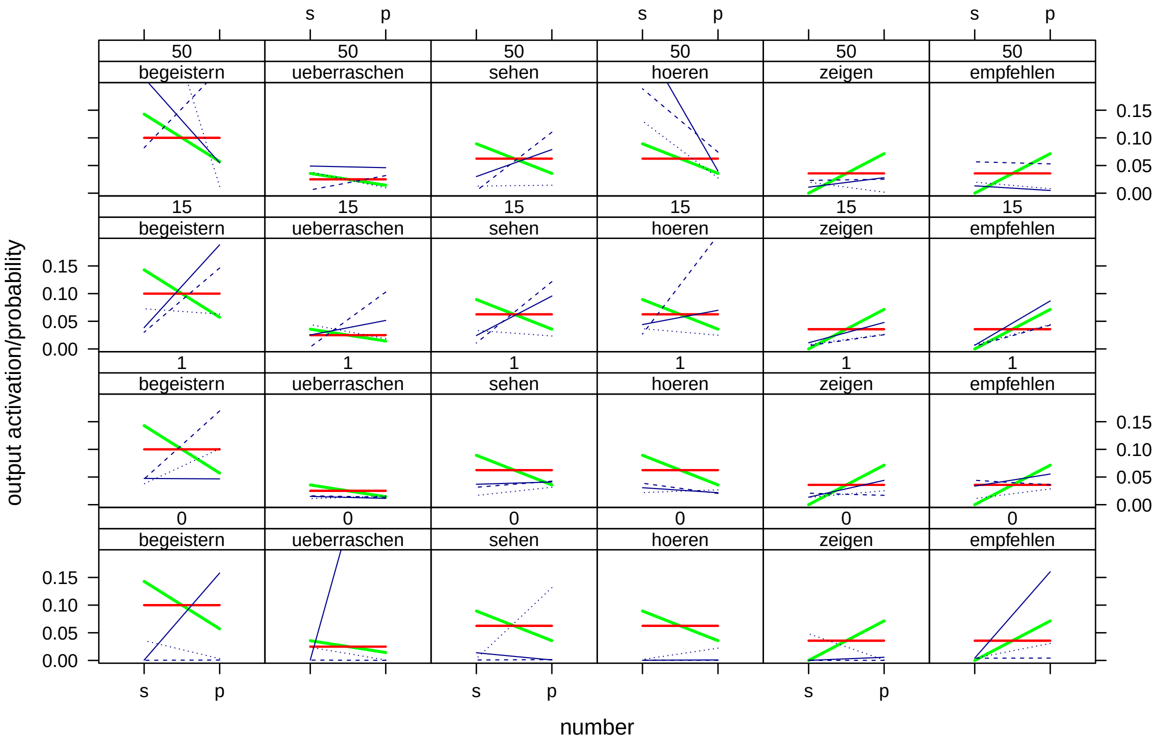


Figure 4.10: Output activation, transitional probability and NI-predictions of verbs at matrix verb position. Output activations of three different networks are plotted. Rows demarcate singular/plural pairs of each of the verbs used in the grammar. This example is one of the rare sentences where the networks do not learn the correct pattern, even after 50 epochs of training. Notice that the direction of the probability slopes (green line) differs for different verb classes, which is due to the temporal ambiguity in the sentence.

method, it turns out that networks also produce a high number of misses in level 2 sentences, and do not seem to make much progress even after 50 epochs of training for level 3 and 4 sentences. Importantly, level 1 sentences, which are the ones we will use for further analysis, are processed mainly correctly, with the exception of some particularly difficult garden-path sentences.

On the one hand, these results clearly show the limits of our model, which can hardly be said to have grasped the recursive nature of relative clauses completely. However, humans also have a hard time when processing multiply embedded sentences, and can hardly grasp the meaning of sentences with more than two central embeddings. Thus, the fact that depth of embedding strongly influences network performance, rather than showing that the model is not suited to the task of language processing, provides evidence that they are particularly plausible one since the results are compatible with experimental evidence.

However, I don't want to make strong claims regarding this issue, since the model was not explicitly designed to answer questions of recursion and depth of embedding. In addition, the grammar incorporates some features that make long distance dependencies especially hard. The inclusion of highly ambiguous elements, which is necessary to investigate local syntactic coherence, inevitably leads to complex garden-path sentences, as is evident in the example presented in section 4.2.2.3.[5] Two phenomena that have been revealed in the analyses have to be kept in mind. On the one hand, ORCs, both accusative and dative, seem to pose more problems than processing SRCs, at least with regard to number agreement at the matrix verb. Although very roughly, this mirrors the complexity difference between SRCs and ORCs for human comprehenders (see section 2.1.2). On the other hand, some cases of temporal ambiguity seem to pose severe problems for the networks. Both phenomena are worth an investigation on their own, and have been modeled with SRNs elsewhere. However, since they are not the main topic of this thesis, I will leave these topics aside for now.

4.3 Effects of Local Syntactic Coherence

In the preceding section I showed that the networks learn the grammar, including long-distance number agreement, to a degree that we can take the model as a plausible candidate for partially modeling human sentence processing. We will now turn to the main question the model aims

[5] The decision not to include the comma further complicates the issue, which can in some cases lead to sentences that can hardly be understood even by a potentially competent language user as the author. Given these aggravating factors, the networks perform surprisingly well.

to address: Do the networks show effects of local syntactic coherence of the kind that has been reported in the visual world experiments?

Effects of local coherence are conceptualized as activations of locally predicted, but globally incorrect nodes. Thus, local coherence effects can be measured as *false alarm* activations of particular nodes at suitable sentence positions. We only focus on cases where the local prediction is induced by more-word sequences, and where the local prediction is incompatible with the globally correct one. To assess these cases, we focus on a particular sentence position where more-word local and sentence-wide global predictions diverge qualitatively. In dative relative clauses like:

(52) der-kommandant-*NOM-SG* dem-*RP-DAT-PL* ...
the-commander-*NOM-SG* whom-*RP-DAT-PL* ...

der-pilot-*NOM-SG* überrascht den-kommandanten-*ACC-SG*
the-pilot-*NOM-SG* surprisedly/surprises the-commander-*ACC-SG*

... zeigt sieht den-piloten-*ACC-SG* EOS
... shows sees the-pilot-*ACC-SG* EOS

the embedded sequence is identical to a simple main clause, which would be continued by an end of sentence marker in the majority of cases:

der-pilot-*NOM-SG* überrascht den-kommandanten-*ACC-SG* → EOS

Globally, however, the correct continuation of the partial sequence is the verb of the relative clause, as in example (52).[6] A local coherence effect should show up as false alarm activation of the end of sentence marker at the position of the RC verb, while no or less EOS-activation should be observed in structurally similar sentences that do not contain locally coherent sequences. We will test this hypothesis by focusing exclusively on the prediction of an end of sentence marker after the locally coherent sequence, and the corresponding sequence in the control conditions.

4.3.1 Materials and Design

Testset *testLOCO* To evaluate effects of local syntactic coherence, we used the subset of *testRC* containing only dative relative-clauses (henceforth *testLOCO*). The sentences are organized so that they included the experimental factors *additional adverb (add.adv)* and *ambiguity of adverb (amb.adv)*, mirroring the design used in the visual word experiments (chapter 3). In

[6]To a lesser degree, relative pronouns are predicted both locally and globally, since any NP can take a relative clause.

addition, in the model we can make use of another control factor: If the subject of the dative relative clause is plural, the *NP-ADV-NP* sequence does not form a locally coherent sequence anymore:

der-kommandant-*NOM-SG* dem-*RP-DAT-PL* ...
the-commander-*NOM-SG* whom-*RP-DAT-PL* ...

* die-piloten-*NOM-PL* überrascht den-kommandanten-*ACC-SG*
* the-pilots-*NOM-PL* surprisedly/surprises the-commander-*ACC-SG*

... zeigen sieht den-piloten-*ACC-SG* EOS
... show sees the-pilot-*ACC-SG* EOS

The inclusion of this factor has the advantage that the length of the sequence ranging from sentence start to the critical element is held constant in target and control items, which is not the case if local coherence is inhibited by adding an additional item. Thus, sentences in *testLOCO* constitute a $2 \times 2 \times 3$ design, incorporating the factors *additional adverb (add.adv)* (yes/no), *number of relative clause (num)* (SG/PL), and *Ambiguity of adverb (amb.adv)* (high-/low/namb). Table 4.5 shows the 12 conditions and one example sentence each.

Table 4.5: Design of *testLOCO*: Sentence types used to evaluate effects of local syntactic coherence in SRN-simulation. Conditions 1 & 2 (*num RC* SG / *add.adv* no/ *ambig.FB*$_{low/high}$) contain the main clause sequence «der-pilot-*NOM-SG* überrascht/begeistert den-kommandanten-*ACC-SG*» and thus a local syntactic coherence

	num RC	*add.adv*	*amb.adv*	example
				der-kommandant-*NOM-SG* dem-*RP-DAT-SG* ...
1	SG	no	ambig.FB$_{high}$	**der-pilot-*NOM-SG* begeistert den-kommandanten-*ACC-SG***
2			ambig.FB$_{low}$	**der-pilot-*NOM-SG* überrascht den-kommandanten-*ACC-SG***
3			non.ambig	der-pilot-*NOM-SG* glücklich den-kommandanten-*ACC-SG*
4		yes	ambig.FB$_{high}$	der-pilot-*NOM-SG* äußerst begeistert den-kommandanten-*ACC-SG*
5			ambig.FB$_{low}$	der-pilot-*NOM-SG* äußerst überrascht den-kommandanten-*ACC-SG*
6			non.ambig	der-pilot-*NOM-SG* äußerst glücklich den-kommandanten-*ACC-SG*
7	PL	no	ambig.FB$_{high}$	die-piloten-*NOM-PL* begeistert den-kommandanten-*ACC-SG*
8			ambig.FB$_{low}$	die-piloten-*NOM-PL* überrascht den-kommandanten-*ACC-SG*
9			non.ambig	die-piloten-*NOM-PL* glücklich den-kommandanten-*ACC-SG*
10		yes	ambig.FB$_{high}$	die-piloten-*NOM-PL* begeistert den-kommandanten-*ACC-SG*
11			ambig.FB$_{low}$	die-piloten-*NOM-PL* überrascht den-kommandanten-*ACC-SG*
12			non.ambig	die-piloten-*NOM-PL* glücklich den-kommandanten-*ACC-SG*
				... **ZEIGT** sieht den-piloten-*ACC-SG* EOS

Local syntactic coherence and its different strength due to the finiteness bias *FB* can be visualized as the n-gram based transitional probability of an end of sentence marker at this position. Figure 4.11 shows transitional probabilities (y-axis) of the EOS at the RC verb position

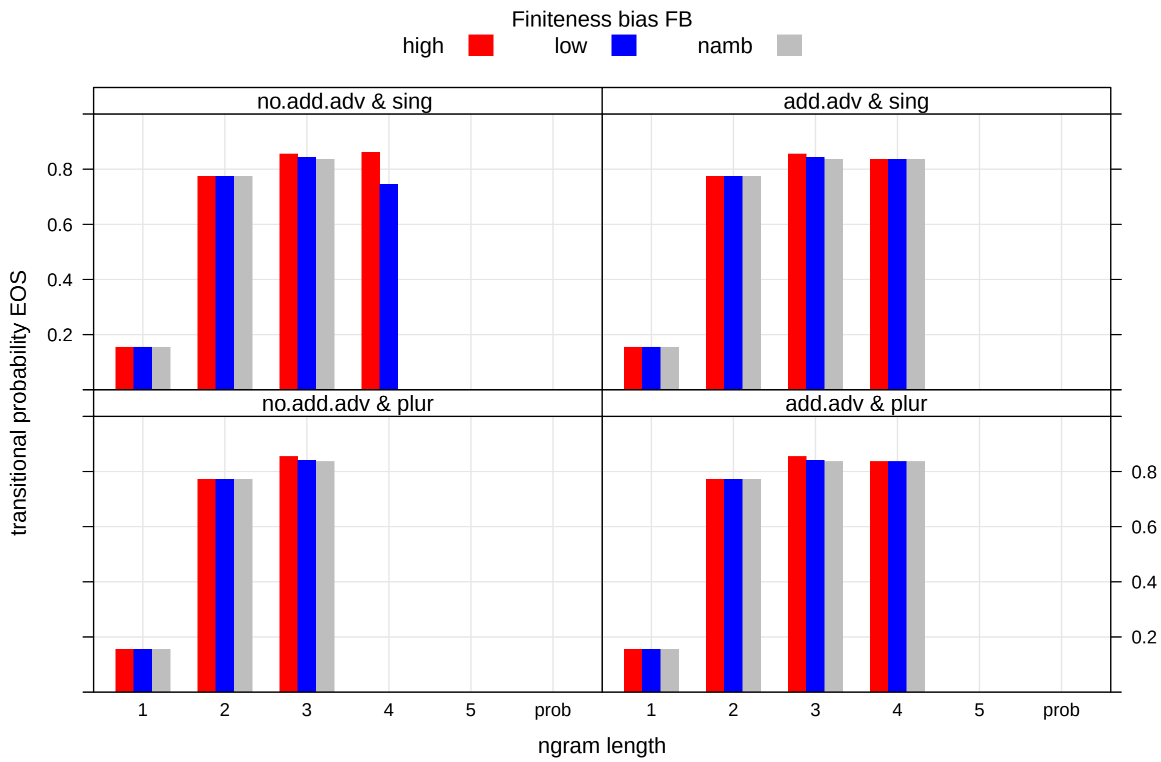


Figure 4.11: N-gram based predictions of the end of sentence marker at RCverb position in dative relative clauses:
e.g. (top left): Der-pilot[prob] dem[ng5] der-kommandant[ng4] überrascht(high)/begeistert(low)/glücklich(namb)[ng3] den-piloten [ng2] –> EOS.
X-axes show n-gram length (1-5) and global transitional probability according to the grammar (prob).

in the different types of dative relative clauses that contain an adverb modifying the RC verb. The figure shows all 12 conditions, according to table 4.5. The groups of three bars each show n-gram based transitional probabilities ($n = 1$ to $n = 5$), as well as the global transitional probability (*prob*) that is derived from the PCFG. Red and blue bars refer to sentences with high and low *FB* of the adverbial element respectively and grey bars stand for control sentences with unambiguous adverbs. As we can see, there's a fairly high 3-gram prediction of an EOS in all conditions, signifying the high frequency of «überrascht/begeistert^3 den-piloten-*ACC-SG*^{2} $\rightarrow$ EOS» (red and blue bars), but also of «glücklich^3 den-piloten-*NOM-SG*^{2} $\rightarrow$ EOS». In both simple main clauses (a) with and (b) without an adverbial phrase, this sequence occurs at the end of a sentence:

(a) der-kommandant-*NOM-SG* [überrascht^3 den-piloten-*ACC-SG*^{2}] $\rightarrow$ EOS

(b) der-kommandant-*NOM-SG* sieht [überrascht/glücklich^3 den-piloten-*ACC-SG*^{2}] $\rightarrow$ EOS

The slight influence of *FB* in 3-gram predictions signifies verb-usage, as in (a).

In the upper left panel we see the crucial difference in EOS prediction between ambiguous and unambiguous adverbs in 4-gram predictions, signifying the presence or absence of a full-clause local syntactic coherence, as well as the smaller difference between high and low *FB*. The high 4-gram predictions for the EOS in sentences with an additional adverb («äußerst»; right-hand panels) are due to the fact that the sequences «äußerst überrascht den-piloten-*ACC-SG*» and «äußerst überrascht die-piloten-*ACC-PL*» will occur most often as the end of an ordinary ditransitive main clause like:

der-pilot-*NOM-SG* zeigt dem-komm.-*DAT-SG* [(äußerst)^4 überrascht^3 den/die-piloten-*ACC-SG/PL*^{2}] $\rightarrow$ **EOS**

and thus predict an EOS. However, since in these cases «überrascht» is used as an adverb, but not as main verb, ambiguity and *FB* show no influence, as is reflected in the equal height of the three bars. The global transitional probabilities of zero in all conditions signify that the EOS is not a correct continuation in any of the conditions – the activation of an EOS at this position is thus a clear false alarm error in all conditions. The values of zero for 4-gram predictions in the left column and 5-gram predictions in all panels reflect that the local syntactic coherence is "disambiguated" when the relative pronoun is taken into account. If the networks show effects of local syntactic coherence, we expect higher false alarm activations of the EOS in the cases with local syntactic coherence, and this effect should be modulated by the Finiteness bias *FB* of the ambiguous element.

Outlier correction Exploration of the data revealed that the data are rather noisy, and some activations seem to be very odd. In a behavioral experiment this would clearly call for outlier correction. A correction for outliers in the classical sense, however, seems to be odd for simulation data, since the reasons for outliers in behavioral data – missing attention, distraction of participants or technical problems – can hardly be taken as legitimation for correction in a connectionist model. However, outliers affect the distribution of data points and thus decrease the quality of the model. In particular, the false alarm activations of the EOS are in general very small, with the consequence that any considerable outlier can strongly distort the results.

One reason for outliers in connectionist networks is that a network gets stuck in a local minimum. Since learning rate was not reduced over training and networks were initialized with a relatively high range of random weights, it may be that some networks fluctuate significantly more than others, in particular in later stages of training. Thus, I decided not to exclude particular data points, but rather to check if individual networks that produce considerably more outliers than others could be identified. If so, it will be helpful to exclude these networks from further analysis to keep noise within a reasonable range. As a first step, I identified outliers separately for each epoch by calculating the mean of raw EOS activation over the whole test-set, and considered data points deviating more than 2.5 standard deviations from the epoch mean as outliers. Figure 4.12 shows the data points considered as outliers for each epoch. The upper panel shows, for each network, a filled box if the network is responsible for one of the outliers in the particular epoch. (In particular the maximal value of about 0.5 in epoch 50 seems very odd). It turned out that three networks (1.4, 3.3 and 5.6, marked in red) produce outliers in almost all epochs. Red dots signify the data points stemming from these three networks in the lower panel. It is clearly visible that these networks produce a large amount of outliers.

4.3.2 Results

Since the activation of the EOS at the relevant RC verb position is ungrammatical, the error and thus the differences between conditions will decrease with training. Possible ways to account for change over training are residualizing EOS activation by training epoch, or to include epochs of training as a further predictor in the statistical model. However, both methods bear the danger that an effect may mainly be driven by strong effects in early stages of training. In the preceding section, however, I showed that it takes a considerable amount of training until the networks have learned the language to a reasonable degree. As a consequence, if effects

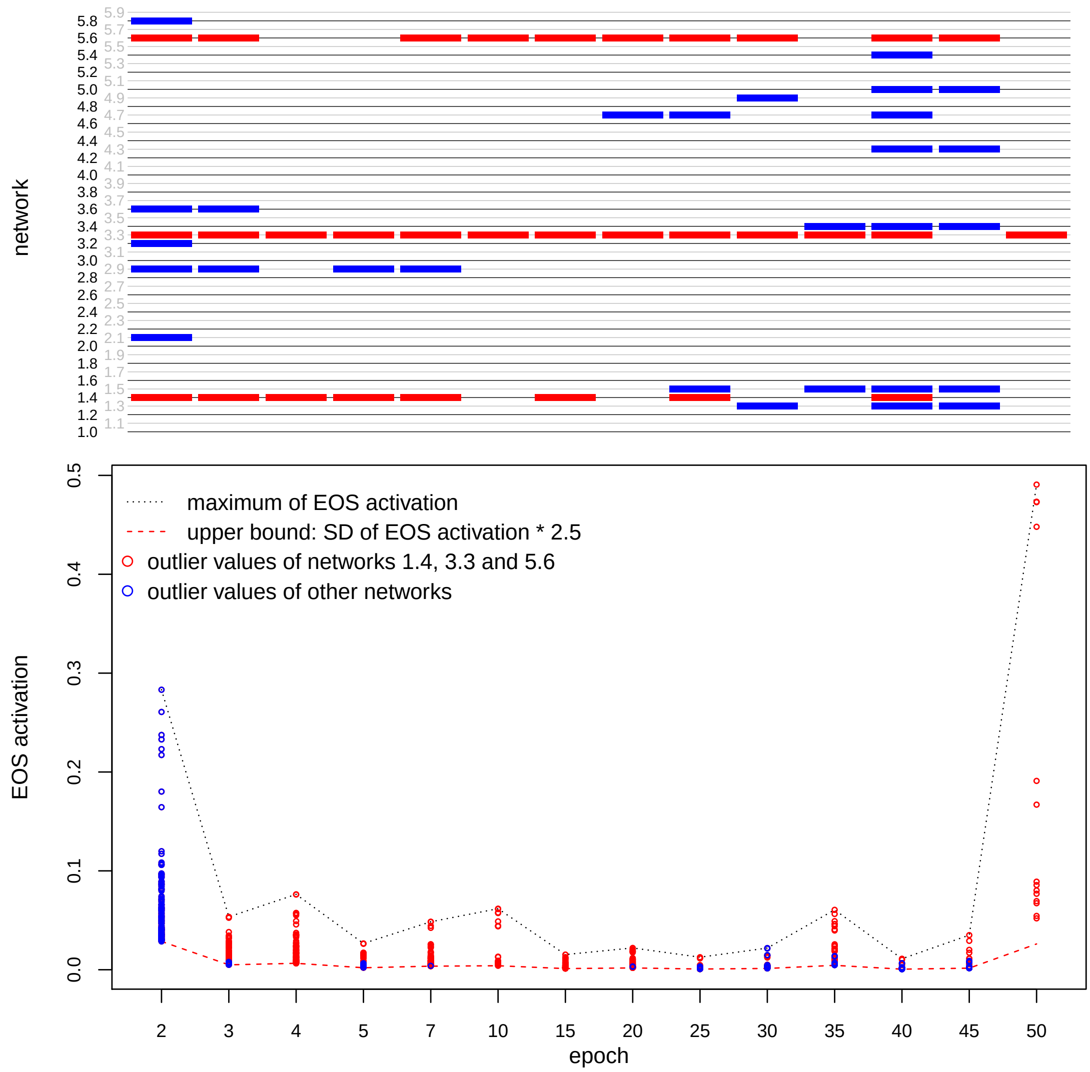


Figure 4.12: Visualization of outlier procedure. Lower panel shows data points (EOS activation at RC verb position) above 2.5 * standard deviation of averaged EOS activation per epoch. Upper section shows which networks produced these outliers. Three networks (1.4, 3.3 and 5.6, marked in red) produced outliers in the majority of epochs and were thus excluded from further analysis. Outlier values produced by the three excluded networks are marked in red in the lower panel.

of local syntactic show up only in early training stages, results are inconclusive with regard to behavior of competent human language users. Thus, we have to ensure that the effects are not only spurious results of insufficient training, but that also “mature” networks exhibit effects of local syntactic coherence.[7]

In a first step, I will visualize and describe the general patterns of EOS-activation and its development over training. In the second I analyze the patterns for statistical validity.

Figure 4.13 shows the means of the logarithmized EOS-activations in a subset of epochs, as predicted by *FB* of the adverbial element (x-axis). Data is grouped by the factors *additional adverb* and *number of RCsubject*. Note that y-axis intercept differs between panels, whereas the scale range (min-max distance) is the same for all panels. As expected, the y-values reveal a continuous decrease of false alarm activation over training, regardless of condition. Considering the patterns between conditions, several observations are noteworthy:

1. In conditions 1 to 6 (1 to 3, *no additional adverb*: solid blue lines; 4 to 6, *with additional adverb*: dashed red lines), we can observe an interaction of *FB* and presence of the additional adverb, such that higher *FB* leads to higher EOS-activation, and more so if the local sequence is not rendered incoherent by an additional adverb. Both the inhibition of a local coherence by use of an unambiguous adverb and by adding an additional adverb leads to lower false alarm activations. In earlier epochs, an effect of *FB* can also be observed in sentences with additional adverb, indicating that the length of local sequences that influence prediction grows over training.

2. Regarding the factor *number of RCsubject*, the picture is less clear. Comparing conditions 1 to 3 (*singular RC subject*: solid blue line) with conditions 7 to 9 (*plural RC subject*: dotted dark green) we can observe a clear effect of *number of RCsubject*: false alarm activation of the EOS is considerably higher if the RC subject is singular. In addition, *FB* also seems to increase EOS activation in the plural cases, although the plural NPs do not constitute local coherence. The effect of *FB*, however, seems to be stronger in the singular sentences.

3. Finally, if both plural and additional adverbs are present, EOS activation is lowest (dot-dashed green line).

[7]This is in particular important because SRN start language acquisition by recognizing local dependencies (Elman, 1993). Effects of local syntactic coherence in a premature stage of learning can thus be viewed as a direct consequence of insufficient training, but not as providing insight into skilled language processing.

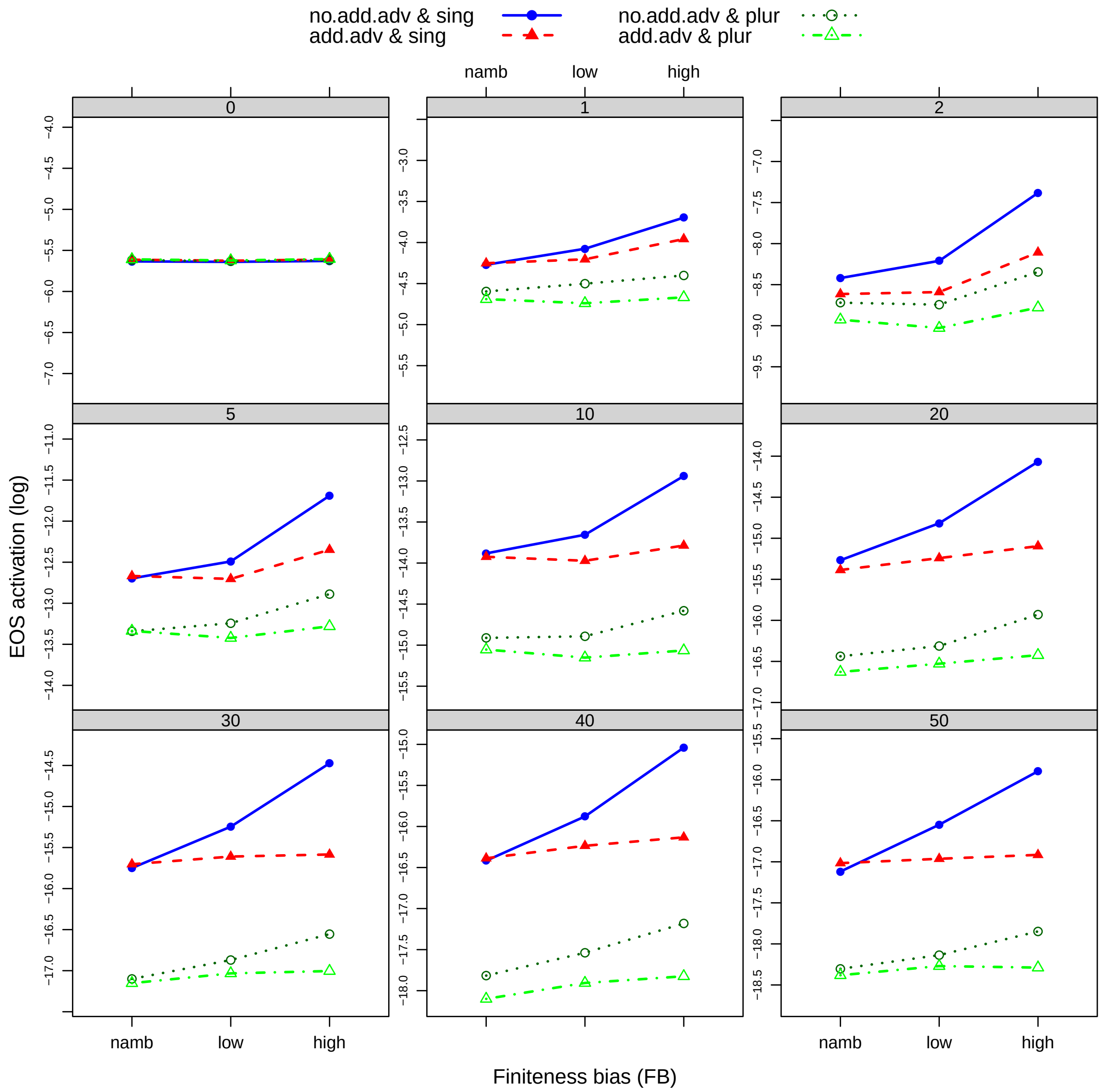


Figure 4.13: Log activation of EOS at RC verb position, averaged over networks. Panels represent different training epochs. Solid blue lines represent conditions 1 to 3, dashed red lines conditions 4 to 6; Green lines represent control conditions with plural RC subject (dotted dark green: conditions 7 to 9; dot-dashed green: conditions 10 to 12. Local syntactic coherence is present in conditions 1 & 2, blue line, right and far right. Y-axis intercept differs between panels, range (min-max distance) is the same for all panels.

Most of the described patterns already emerge in early stages of training, but stay stable even after 50 epochs of training. Thus the patterns are consistent with the hypotheses formulated above. However, it is also evident that the false alarm activations do not simply match the n-gram-based local transitional probabilities visualized in figure 4.11. N-gram models only predict an effect of *FB* for conditions 1 to 3. In particular, the considerably higher EOS activations in sentences with a singular RC subject are not expected.

In the next two sections, I will consider the data as implementing two separate experiments, *SRN experiment 1* and *SRN experiment 2*. The first one strongly resembles the design of visual world experiment 2 (section 3.4) and incorporates the factors *additional adverb* and *FB*. The second one includes *number of RC subject* and *FB*. Splitting the data reduces the complexity of the statistical model by reducing the number of factors that have to be considered. In addition, *SRN experiment 1* closely matches the design of visual-world experiment 2, while data is not provided with regard to *number of RC subject*.

4.3.2.1 SRN Experiment 1: *FB* × Additional Adverb

Table 4.6: Designmatrix of SRN experiment 1. Numbers refer to table 4.5. Conditions containing a locally coherent sequence are marked in boldface. Reference cell for the statistical models (intercept) is the “maximally incoherent” condition (6)

SRN Exp. 1		*additional adverb*	
		no	yes
FB	high	**(1)**	(4)
	low	**(2)**	(5)
	namb	(3)	(6) **Intcpt**

SRN experiment 1 includes the factors *FB* (not ambiguous / low *FB* / high *FB*) and *additional adverb* (yes / no) and thus incorporates conditions 1 to 6, according to table 4.5. The test set consists of 16 sentences per condition, varying in number of the matrix clause, number of the accusative object and the particular nouns and verbs used in the sentence.

Hypothesis In the analyses, the factor ambiguity is conceptualized as *FB* comprising three levels, considering unambiguous adverbs as having a *FB* of 0. Effects of local syntactic coherence should result in an interaction of *FB* and an additional adverb such that false alarm EOS activation will be higher in sentences without an additional adverb and modulated by *FB*.

Statistical model EOS activation drops significantly during learning and therefore effect sizes will equally decrease with number of training epochs. To evaluate if effects on false alarm prediction of the EOS stay stable over training, separate linear mixed effect models were fitted to the data of different training epochs.[8] Since the distribution of data points is strongly skewed to the right, EOS-activation is logarithmized for further analysis, thus providing a normal distribution of data points.

The models include logarithmized EOS activation at the RC verb position as the dependent variable, and the fixed factors *FB* and *additional adverb*. To account for differences between individual networks and for variations between the test sentences, random intercepts for network, number of matrix sentence (SG/PL) and number of accusative object (SG/PL) are included.

Factors are treatment coded. As reference levels, *namb* is selected for the factor *FB*, and *yes* for the factor *additional adverb*. Thus, any effect of local coherence will result in positive estimates and *t*-values, signifying an increase of EOS-activation compared to the reference levels *FB:namb/add.adv:yes*. For each epoch, *t*-values and *p*-values are calculated, where *p*-values are derived by MCMC sampling (number of simulations = 5000), using the *pvals.fnc* provided in the *langaugeR* package (Baayen, 2009).

Figure 4.14 shows, in the upper panel, the *t*-values derived from the models fitted to the data separately for each epoch. *T*-values greater than 2 (grey horizontal line) signify reliable effects. Solid lines represent interaction of *Additional adverb* × low (red) and high *FB* (blue). Dashed lines represent simple contrast of the factor levels and intercept. To facilitate interpretation and give an impression of the effect size, the bottom panel represents the log EOS activation, averaged over all networks, after 2,10 and 30 epochs of training. With regard to performance on number agreement, networks after 30 epochs of training can be considered as "mature".

The models revealed that *FB* interacts significantly with *additional adverb*, for both high and low *FB*, already in early stages of training. Although false alarm activation drops to values near 0 during training, the interactions are significant even after 50 epochs of training. As can be seen in the lower left plot in figure 4.14 (epoch 2), at the beginning of training EOS activation is in general considerably higher for high *FB*, indicating that the networks seem to group unambiguous adverbs and "almost unambiguous" adverbs (low *FB*) together.

[8] For pragmatical reasons, I will include only a subset of epochs in the analysis. Epoch 0, serving as a baseline indicating what the randomly initialized networks predict, and epochs 1,2,3,5,7,10,15,20,25,30,25,40,45,50 (end of training).

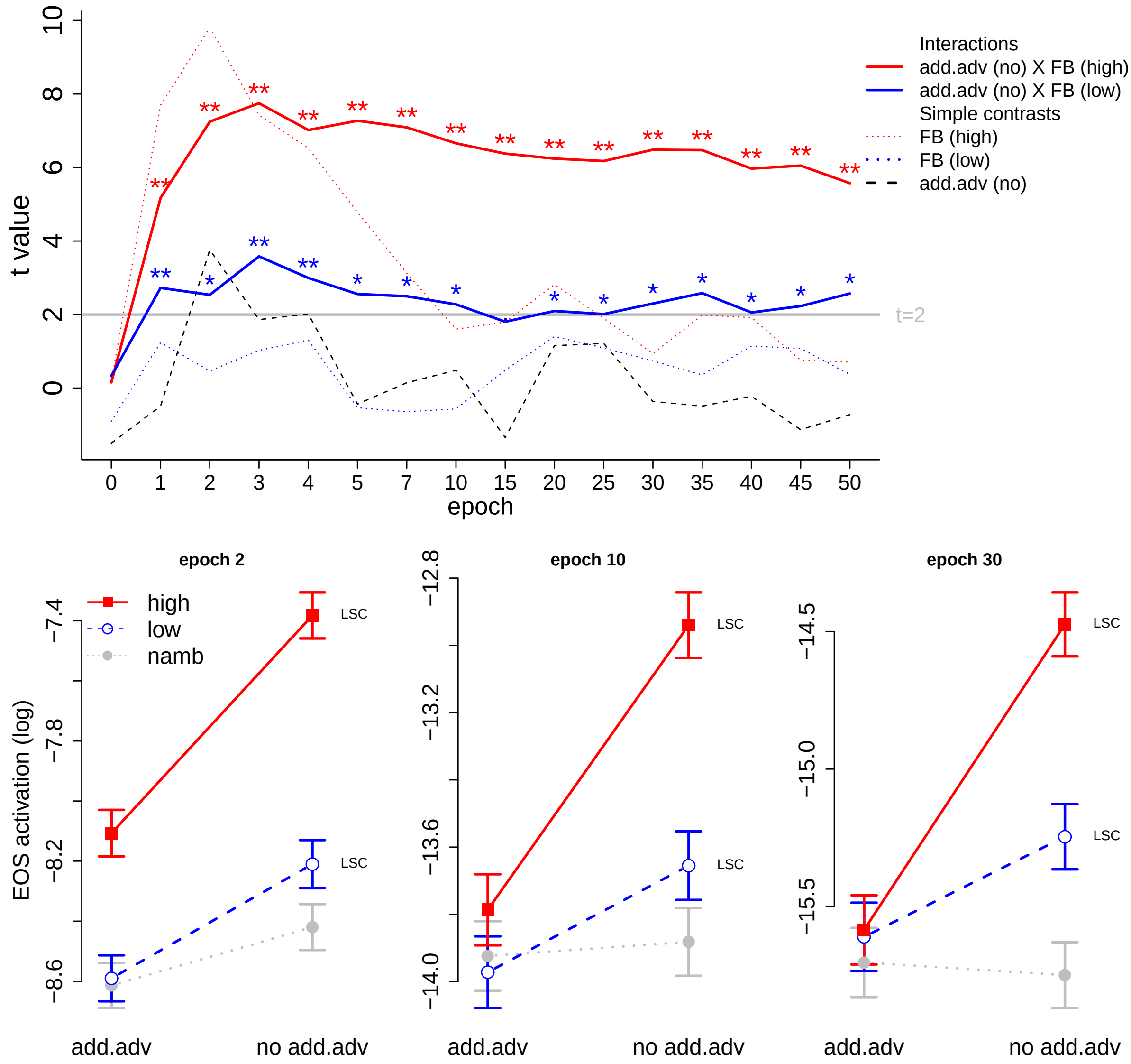


Figure 4.14: Results SRN Experiment 1. Upper panel: *t*-values of interactions and simple contrasts. Lower panels: averaged EOS activation (log) in epochs 2, 10, and 30.

In addition, the higher EOS activations in sentences without an additional adverb, even for unambiguous items, signifies that unambiguous elements are regarded as similar to ambiguous ones, thus leading to incorrect EOS predictions although they are not even consistent with the local context. Both the influence of an additional adverb on unambiguous elements, as well as the influence of *FB* in sentences with an additional adverb vanish after sufficient training. The effects of local coherence, however, as well as the modulation of the effect by the bias of the ambiguous elements, stay stable until the end of training, in accordance with the hypothesis.

4.3.2.2 SRN Experiment 2: *FB* × Number RC subject

Table 4.7: Designmatrix of SRN experiment 2. Numbers refer to table 4.5. Conditions containing a locally coherent sequence are marked in boldface. Reference cell for the statistical models (intercept) is the "maximally incoherent" condition (9)

SRN Exp. 2		*number of RC subject (num)*	
		singular	plural
	high	**(1)**	(7)
FB	low	**(2)**	(8)
	namb	(3)	(9) **Intcpt**

In addition to the additional adverb as a control factor, I included number of RC subject as a further factor in testset *testLOCO*. As has been mentioned above, in contrast to adding an additional element as in SRN experiment 1, manipulating number of the RC subject does not change the number of elements that have been encountered within the sentence before reaching the RC verb position. However, just as with the additional adverb, a plural RC subject does inhibit local syntactic coherence. Thus, SRN experiment 2 includes the factors *FB* (not ambiguous / low *FB* / high *FB*) and *number of RC subject* (singular / plural), incorporating conditions 1 to 3 and 7 to 9, according to table 4.5. The test set consists of 16 sentences per condition that vary in number of the matrix clause, number of the accusative object and the particular nouns and verbs used in the sentence.

Hypothesis Effects of local syntactic coherence should result in an interaction of *FB* and number of RC subject (*num*) such that false alarm EOS activation will be higher in sentences with singular RC subjects, modulated by *FB*.

Statistical model SRN experiment 2 was analyzed in the same way as SRN experiment 1 by fitting separate linear mixed effects models to the output data of different epochs. The models include logarithmized EOS activation at the RC verb position as the dependent variable and the fixed factors *FB* and *number of RC subject*. Random intercepts for network, number of matrix sentence (SG/PL) and number of accusative object (SG/PL) constitute the random effect structure.

As reference levels, level *namb* was selected for the factor *FB*, and *plural* RC subject for the factor *num RC*. Thus, any effect of local coherence will result in positive estimates and *t* values, signifying an increase of EOS-activation compared to the reference levels *namb/plural*. For each epoch, *t* values and *p* values are calculated, where *p* values are derived by MCMC sampling (number of simulations = 5000), using the *pvals.fnc* provided in the *langaugeR* package (Baayen, 2009).

Figure 4.15 shows, as in the preceding section, *t*-values in the upper panel, and log EOS activation, averaged over all networks, after 2, 10 and 30 epochs of training. In contrast to SRN experiment 1, only high *FB* reveals a significant interaction with the *num RC* over the whole range of epochs. As is visible in the lower panels, high *FB* leads to considerably higher EOS activation in both singular and plural conditions. Most evident, however, is the strong effect of *number of RC subject* on EOS activation such that the networks activate the EOS significantly more if the subject of the RC is in singular, regardless of *FB*.

This effect reveals a general problem that has to be considered when evaluating activation of one particular output node. The training algorithm used here (and in the majority of SRN models) ensures that output activation of all nodes sums up to 1. With regard to the activation of a particular node this has the consequence that its actual activation value depends on all other output values at the particular position. Thus, any differences between conditions will not only result in different activation of that particular node, but also in a variety of other effects. Since in experiment 2 *number of RC subject* serves as an experimental factor, any differences with regard to performance in number agreement on the RC verb will inevitably result in systematic differences of output activation, in particular of verbs. If, for example, correct singular verbs at the RC verb position are predicted less accurately than plural verbs, the pattern of false alarm activation of other nodes, including the EOS, will be influenced. Differential difficulty of number agreement for singular and plural sentences can, for example, result from the fact that only singular verbs in the grammar are ambiguous, but none of the plural verbs. However, a more sophisticated analysis of this rather complex issue is beyond the scope of this work.

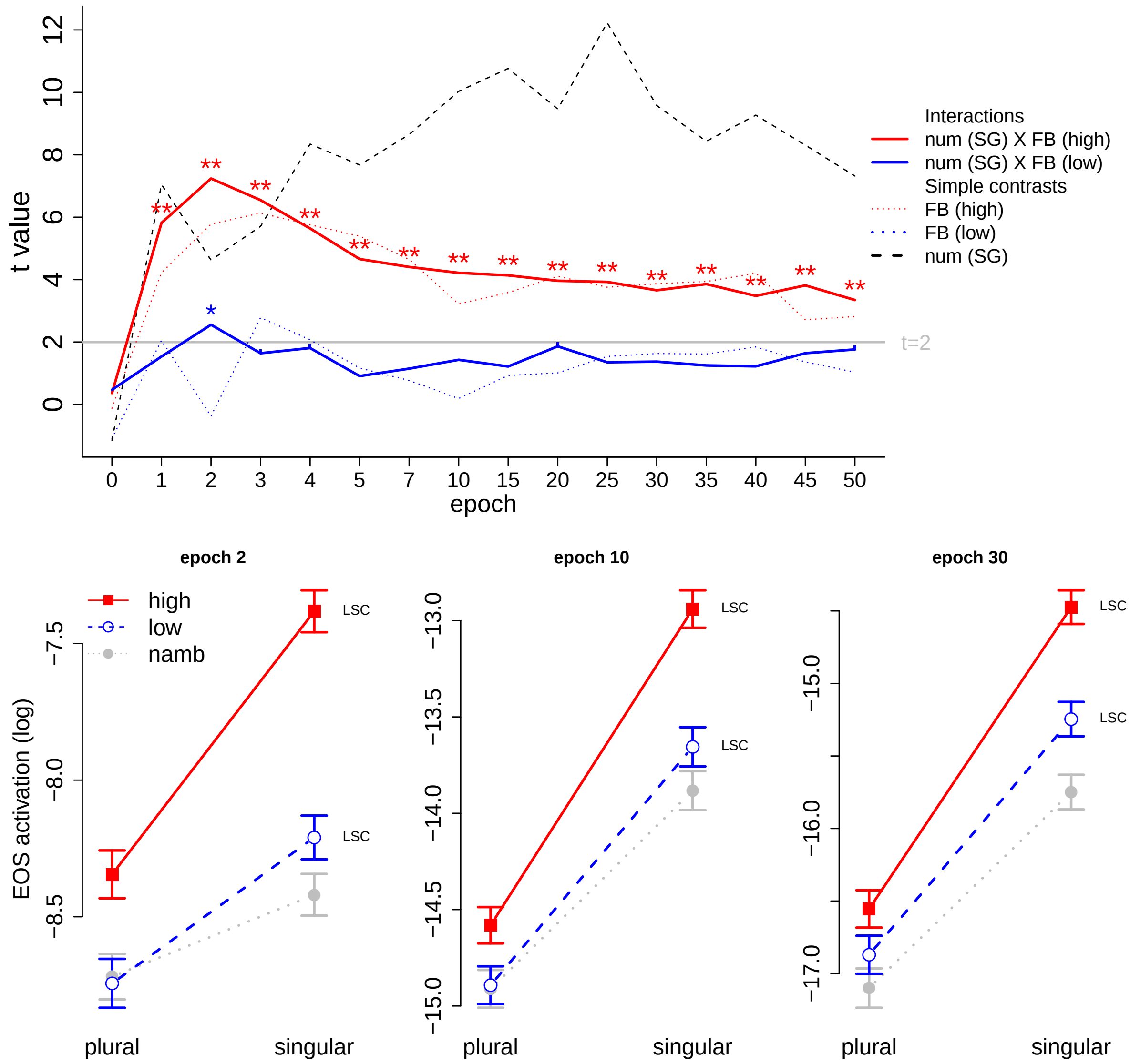


Figure 4.15: Results SRN Experiment 2. Upper panel: *t*-values of interactions and contrasts. Lower panels: averaged EOS activation (log) in epochs 2, 10, and 30.

4.3.3 Preliminary Discussion

In the preceding section I evaluated the hypothesis that SRN models that have learnt a reasonably complex language elicit effects of local syntactic coherence. Effects of local syntactic coherence were measured as differences in false alarm activations of the end of sentence marker at the RC verb position, where globally correct predictions and only locally coherent predictions maximally diverge. The hypothesis has been evaluated by conducting two experiments using different control conditions.

In SRN experiment 1, I used test sentences derived from the design of visual word experiment 2 (section 3.4), while SRN experiment 2 included a further factor which was not implemented in the visual world experiments. The results of both experiments revealed that in sentences containing a full-clause local coherence the networks activate the EOS more strongly than in a variety of control sentences where local coherence was inhibited either by using unambiguous adverbs, an additional modifying adverb or changing the number of the relative clause's subject. Statistical analysis revealed that ambiguity interacts with presence/absence of an additional adverb, and, although less clear, with number of RC subject, revealing that false alarm activation of the EOS is not induced by the ambiguous element alone, but that a larger locally coherent context is necessary to interfere considerably with the globally correct predictions.

In addition, effects of local syntactic coherence are modulated by *FB* of the ambiguous element within the locally coherent sequence, indicating that the networks represent ambiguity of an element not as a binary category, but as a graded property of an element. Both the effects of local coherence as well as influence of *FB* are in accordance with the results of visual world experiment 2, which revealed similar effects of both ambiguity and *FB*.

The analysis of the SRN experiments also revealed effects that can not readily be linked to available behavioral data, but that are of interest for two reasons. On the one hand, further investigation of these effects can foster a better understanding of SRNs and how they do what they do. On the other hand, taking the model as a generative cognitive theory, it will be interesting to evaluate similar effects in further behavioral experiments.

5 General Discussion

In this chapter, I will point out some potential issues and open questions, mainly with regard to the corpus analysis (section 3.5) and the SRN model (chapter 4).

5.1 Local Coherence Revisited

In section 3.5, I showed that the finiteness bias *FB* determines if and how strong a locally coherent sequence influences gaze behavior. However, this operationalization of local coherence as a graded, quantitative property is, of course, only one of many possible ways of measuring this. In addition, it can obviously not be applied to cases where local syntactic coherence does not depend on lexically ambiguous items, as, for example, in the visual world experiments evaluating local coherence effects on anaphora resolution (Weldle, 2011, see section 2.3.6.1) or Konieczny, Müller, Baumann, et al. (2009). Stronger support for the hypothesis that effects of local syntactic coherence are a result of frequently encountering a particular sequence in other contexts would be provided by a more elaborate analysis, taking into account not only the lexical category bias, but quantifying coherence of the entire sequence. So far, a first attempt to evaluate n-gram properties of the stimuli used in visual world experiment 2 was dismissed because no statistically useful amount of occurrences of the relevant elements was accessible. As described in section 3.5, even analyzing only the ambiguous items accurately turned out to be rather difficult. Although considerably large corpora are available, the main problem was to derive reliable category frequencies for the ambiguous elements. Thus, more work has to be done regarding this question. However, since both the availability of large corpora, as well as the reliability of annotations constantly improves, a range of different measures will hopefully complement the approach adopted here.

In addition to effects of frequency, a number of other factors are likely to determine if, or how strong, a local sequence will distract attention, or influence processing complexity. Tabor et al. (2004) already provided evidence that semantic fit within the local syntactic coherence plays a role. In an eye-tracking while reading study, Müller, Hachmann, and Konieczny (2009) provided further evidence that local *syntactic* coherence alone does not suffice to induce effects on processing, but that semantic incoherence within the local sequence inhibits the effects. Currently, we are running an fMRI study where we are investigating if argument properties that induce semantic violations *only within* the locally coherent sequence elicit brain responses similar to the ones known to respond to semantic violations in sentences.

Another open question is whether the length of a locally coherent sequence can fortify interference. This would be akin to "digging in effects" that have been reported with regard to ambiguous sentences in Tabor and Hutchins (2004), showing that recovering from a garden path gets more difficult if information within the ambiguous region supports the initial, incorrect path. Can a particularly long and strong coherent sequence even force a comprehender to loose track of the global structure and settle in at the local one?

5.2 Modeling Effects of Local Coherence with SRNs

The simple recurrent network model proposed here provides only a very limited picture of the complex processes and mechanisms underlying human language processing, and thus a number of critical issues can be raised. These issues can be roughly grouped into considerations about input and output. Of course, the architecture used as such can be criticized as well. However, I will not discuss the general suitability of SRNs with regard to, for example, neural plausibility, or the very narrow approach of the task of language acquisition inherent in the prediction task. This has been done in length elsewhere. However, some aspects of the particular model proposed here are worth discussing. Concerning the input, the main questions regard the grammar and thus the materials that were used to train the network.

5.2.1 Input: Training Materials

The grammar used to generate the sentences the networks are trained on is a rather crude simplification of any natural human language. In some sense, it can be said to be both much too simple and, at the same time, unnecessary complex. In other words, the underlying language model does not provide a very close fit to the linguistic environment human language users are confronted with and adapt to.

Regarding simplicity, the grammar obviously deviates from natural language in several aspects. The artificial language includes only a very limited set of word classes and subclasses of categories, as well as a low number of instances for each class. The set of possible sentence structures only represents a tiny subset of the possible structures of German. The grammar does not include any semantic differences (beyond a whiff of verb semantics), as, for example, animacy or differences in co-occurrence frequency between particular nouns and verbs that are known to have a considerable influence on processing. The explicit case and number marking as well as the representation of determiner-noun phrases as one node reduces potential complexity considerably. Items that are not distinguishable by form in natural language are represented as orthogonal vectors, which obscures potentially important properties of a language and induces considerable differences in distributional properties of surface forms. In addition, I did not attempt to provide a close match of the probabilities of the grammar with natural language. Put simply, the claim is that the simplified language is not close enough to natural language to support the hypothesis that the model tells us anything about natural language processing. Admittedly, it is hard to argue against this broad claim, since it can be raised against virtually any cognitive model in one way or the other.

However, the aim of the model was not to provide a broad coverage model of sentence processing in order to predict processing costs or reading times. This has been done with a considerably more complex architecture and language in Rohde (2002), and, using a subset of a natural language corpus, in Frank (2009, 2013) and other approaches. Rather, the aim was to show that simple recurrent networks provide a rather parsimonious framework that can account for particular behavioral results – effects of local syntactic coherence – without positing any additional modules or mechanisms in addition to the ones that provide the means to acquire and process a language. However, to evaluate the claim that it *does* in fact acquire this language, the complexity of the language had to be kept within reasonable bounds. This included, on the one hand, keeping the set of possible syntactic structures small enough to be able to analyze performance adequately, which, as it turned out, is not easily done even with a relatively simple language. On the other hand, matching frequencies with natural language would have added considerable complexity, rendering the identification of relevant influential factors virtually impossible.[1] To conclude, facing the trade off between "naturalness" and simplicity, I have chosen the way of proposing a small scale model that, however, includes the relevant properties to investigate the questions at hand by (a) allowing to implement the design

[1]In addition, matching probabilities of a simplified grammar includes a great amount of rather arbitrary decisions about which structures are to be counted and, crucially, at which level of granularity frequencies are to be matched. Thus, the "naturalness" of the resulting grammar will entail a fair amount of assumptions about the "relevant" dimensions of frequency.

of the visual world experiment and (b) provide enough complexity to evaluate if the networks can grasp regularities that are crucial properties of natural languages. Still, the drawback is that we cannot take it for granted that the model can be easily scaled up to a more realistic linguistic environment.

As mentioned above, the grammar can also be regarded as too complex. In particular, the need to include ambiguous elements with different properties and different (transitive and ditransitive) verb classes leads to a complex pattern of transitional probabilities, as has been shown with regard to the supposedly simple sentence (49) discussed in 4.2.2. It turned out to be impossible to process for the networks, due to the complex interaction of ambiguity, embedding and the resulting pattern of verb probabilities. This rather unexpected result supports the claim that it is warranted to keep the grammar even simpler to provide analyzable results.

A potential issue with this regard can also be seen in the decision not to include commas in the grammar. Although this was warranted insofar as the behavioral data to be modeled relies on auditory presentation, it is also clear that clause boundaries are typically marked by prosodic cues, which are thus not mirrored in the training materials of the model. However, it is at least questionable if commas – or, for that matter, any boundary markers – should be represented as having the same status as other lexical elements, which would have been necessary in order to include them within the localist representation of input and output. However, the decision inevitably adds a fair amount of difficulty to the task to achieve long distance dependencies, in particular in combination with ambiguity. This issue was exemplified by example (49): the garden path would have been disambiguated at the comma, and thus predicting the correct number at the matrix verb should have been much easier. With regard to this issue, it is difficult to take a clear stance. At least, it can hardly be argued that omission of commas makes the task overly easy – instead, the fact that the models did achieve number agreement to a reasonable degree without the help of explicit boundary cues provides further evidence for the capabilities of the architecture.

The localist mode of representation of input (and output) is another potential issue. Firstly, it is just a very simple and easy way to go, and it is not easy to tell if it adds or reduces complexity overall. On the one hand, the representation of lexical items as orthogonal bit vectors abstracts from all information that is encoded in the patterns and similarities between different words, which can obviously not be used by the network if similarity is not represented. On the other hand, the model would have to comprise the additional task of sorting out which surface similarities are relevant, and thus have to be represented internally to achieve good performance. Regardless of the question of higher or lower difficulty, the more elaborated the representational format, the more assumptions have to be made about which information is accessible to the system.[2] Thus, although the localist mode of representation constitutes a considerable deviation from the properties of natural languages, it provides a reasonable and simple way to evaluate how far we can get with sequential information alone (Elman, 1990) and thus seems appropriate for the aims of the model with regard to the main questions that were at stake here. A variety of different mappings have been proposed and implemented in other SRN models (e.g. Frank, 2013) providing both the means to use larger corpora [3] and to incorporate a variety of potentially relevant factors that could also be interesting with regard to effects of local syntactic coherence.

An interesting aspect of this debate is that the question of what makes the task harder or easier is not as easily answered as it may seem. For example, although ambiguity induces considerable difficulty on the first view, the possibility of using an item in more than one function considerably reduces the number of items that have to be learned. Similarly, including a greater variety of sentence structures does not necessarily add complexity, but can also provide richer information that can support appropriate generalization. However, this issue is beyond the scope of the current work.

5.2.2 Output 1: Interdependency of Output Values

To investigate effects of local coherence, I analyzed activations of one particular output node, the EOS. Although this decision is well motivated by the question at hand, the method also revealed a problematic aspect. Since the training method enforces output activations to add up to one, the value of a particular node crucially depends on the activations of other nodes. In our experiments, differences in the test sentences lead to differences in EOS activation that are, at least in some cases, much stronger than the effects of local syntactic coherence (section 4.3.2.2). These effects cannot be simply attributed to local transitional probabilities on the surface level, which did not predict these effects.

In addition, it turned out that the number of noun phrase preceding the RC verb position has a considerable influence on EOS prediction – or, rather, on prediction of verbs which is interlinked with EOS prediction. An obvious way to eliminate this source of complexity in the

[2] Of course, representing words as basic units also implements a relatively strong commitment to the word as a basic linguistic unit.

[3] The number of terminal nodes is, for pragmatic reasons, rather limited with a localist input representation: since each terminal node requires one input and one output node, and number of connections grows with each node, training with a realistically large lexicon is hardly feasible.

data would have been to use a more homogeneous test set. However, it seems more valuable to include variation in the test set to support the claim that the effects are a general property. At the same time, it makes clear that averaging over different sentence types bears the danger that the effects are mainly driven by a particular type, and thus may not readily reflect the proposed influence, but rather something which has not been considered at all. This problem, of course, is known from behavioral experiments. Using a presumably simple grammar and a rather simple computational device may suggest that this danger is banned, since all potential sources of noise can be controlled for. However, as we saw, it is not always easy to infer the complex pattern of transitional probabilities and, crucially, the influences of these properties on the internal representations of the network and the resulting output.

On the other hand, to detect interesting differences and generate unexpected results is also a strong advantage of the approach adopted here. Connectionist models, in some sense, do not worry about the rules of the grammar, or the aim of the modeler. They use whatever information is available in the data, no matter if this information was intended or considered irrelevant. While this can be rather frustrating from time to time, it can, on the other hand, tell the researcher quite a bit about influences and patterns she would probably never have thought of. However, this also entails that it is necessary to closely listen to what the networks tell. I tried to do this, in particular with regard to number agreement. However, a lot of effects are still puzzling, and further analysis will tell which ones are merely due to the particular grammar and implementation, and which ones may have the potential to reveal insights into human behavior.

5.2.3 Output 2: low EOS False Alarm Activations

A further potential criticism regards the very low values of EOS activation that the reported effects are based on. Since they are false alarms and thus errors, low values are an inevitable consequence of successful learning, and successful learning is, as I argued repeatedly, an important precondition if the model should be taken as a cognitive model and not merely as a device to calculate a particular input-output function that partly describes behavioral data (or, for that matter, is proposed as a specialized preprocessing module). However, it may be claimed that the resulting low error values can hardly be argued to account for measurable differences in reading times or even in eye movements in a visual world experiment. However, it should be noted that I did not argue for any particular quantitative relation between false alarm activation and behavioral measures. On the one hand, the model is far too simple to make any claims beyond qualitative similarity between patterns in network behavior and human data. On the other hand, the actual values are dependent on the particular parameters and algorithms chosen to train the networks. Although a range of different parameter sets was tested in preliminary simulations, the parameters were not chosen to fit behavioral data as closely as possible, but rather to enable the networks to learn the grammar to a reasonable degree. Presumably, it would have been rather easy to produce much stronger effects, either by reducing the percentage of relative clauses, using a lower learning rate or by just inspecting earlier epochs. Thus, the approach taken here is rather conservative with regard to the evaluated effects.

In general, it should be clear that the model, rather than being a detailed model of human sentence comprehension, has to be merely thought of an instantiation of a class of devices that provide the basis for more detailed models. Crucially, however, the results show that this instantiation provides the means to account for particular effects in human behavior, and does so on the basis of the very same mechanisms that also enable it to learn a grammar from positive input alone. Although several models attempt to account for effects of local syntactic coherence, I take connectionist models of the kind presented here as a particularly good starting point. SRNs provide a promising candidate to account for the general influence of distributional properties in the environment of an organism and the influence of the implicit learning process of these properties on processing.

5.2.4 The Link Between Output and Behavioral Data

The fact that the model does not provide a very close link to the behavioral data from the visual world experiments has already been acknowledged. With regard to previous results which have mainly been observed in reading data, the matter is somewhat difficult. The operationalization of local coherence effects in the model as false-alarm prediction after the locally coherent sequence has been chosen to measure the effect of the embedded main clause sequence on processing. Given the rather indirect link between model output and visual world data, this method has been chosen to mirror the processing of the local coherent sequence as a whole, which I argue is necessary to explain the fixation patterns in the visual world experiments. However, the modeling results can thus not be linked directly to existing results where effects of local syntactic coherence have been observed as elevated reading time *during* the local syntactic coherence, as in Tabor et al. (2004) and Bicknell et al. (2009).

Competition, facilitation and intermediate states of mind Tabor et al. (2004), as well as the majority of other accounts, propose that effects of local syntactic coherence are a result of interference or competition between local and global parses. Since the aim of the SRN model has not been to model processing time, it does not readily provide a mechanism that implements competition in the strict sense, nor explicitly represents global or local (sub)parses. Thus, it does not incorporate the process of settling into a stable state, like the SOPARSE model in Tabor et al. (2004), a time-consuming process of resource allocation, as it is proposed in Bicknell et al. (2009), or processes of reanalysis as in Crocker and Corley (2002).

Since the SRN models used here operate on the string level, and do not strictly "assign" part-of-speech tags to input elements or build up a syntactic structure, it would, however, rather be expected that a strong, local prediction should *facilitate* processing *as long as local and global predictions overlap*, as is the case during processing of the locally coherent sequence.

This point has already been discussed in Levy (2008a, pp. 40) with regard to local syntactic coherence. As has been mentioned before, syntactic surprisal cannot account for effects of local syntactic coherence straightaway. As Levy argues, weakening the constraints that define the well-formedness of trees, thus allowing for a kind of "local surprisal", would result in easier processing during the locally coherent sequence, similar to results regarding facilitative ambiguity. However, reading data so far does not support facilitation by local syntactic coherence, but shows higher processing costs. This problem can thus also be regarded as an issue for the SRN approach adopted here.[4]

However, there are two points have to be taken into consideration. On the one hand, reading as a self-paced task may be more compatible with the notion of settling into a stable state or reallocation of resources at a particular word before proceeding to the next. When processing spoken language, however, the next element arrives regardless of whether the system has accomplished a congruent structural analysis or not. Thus, it is not easy to derive predictions with regard to the visual world data, since facilitation does not readily translate into clear predictions about the resulting fixation patterns on corresponding scenes.

A tentative hypothesis would be that until conflicting evidence arises, as at the end of the local syntactic coherence, all potentially matching objects or scenes attract some degree of attention as long as they are compatible with the concurrent evidence of local and global context. This hypothesis is, as far as I can see, implemented in the impulse processing frame-

[4] However, there is also some evidence for facilitation during local syntactic coherence in reading. In an eye-tracking while reading study, Müller et al. (2009) reported lower first-pass reading times *within* a locally coherent region, but elevated regression-path durations *after* the locally coherent sequence ended.

work (Kukona & Tabor, 2011), which does not operate on competition and settling, but rather describes the continuous time course of the system state as influenced by several sources. Inherent in that proposal, and also consistent with the view adopted here, the system is not always in a clear state of mind with regard to the input that has been processed, but, as has been nicely put in Spivey (2007, p. 182) with regard to syntactic ambiguity:

> [. . .] partially active misinterpretations of a sentence can slow down comprehension merely by causing the system to linger in nameless intermediate regions of state space, in between the two syntactically permissible versions of the sentence.

With regard to local syntactic coherence and the visual word data, local and global coherence would cause the system to linger in between attractors built up by local and global context, which leads, through a stochastic process, to a higher probability of fixating the local scene, compared with cases where the local attractor does not build up.

In the modeling approach adopted here, the focus was not on an elaborated account of the mechanisms that integrate visual and linguistic input, and thus does not provide a cognitive model of the visual world results in the strict sense. In contrast, I proposed SRNs as providing a mechanism that uses distributional properties in the linguistic environment to achieve a level of performance that resembles human behavior in important respects, including phenomena that are hard to account for by assuming a grammar-based rational parsing mechanism. Thus, SRNs can serve as a plausible candidate that provides the underlying mechanisms further processes can be built on.

5.2.5 SRNs as Generative Theories

Finally, the best way to evaluate whether a cognitive model is a plausible candidate to explain human behavior appropriately is to test its distinctive predictions in behavioral experiments. The systematic false alarm activation of locally predicted, but ungrammatical elements has been one such prediction. In the meantime, empirical results have provided strong evidence that these effects do not indicate the inadequateness of SRN models, but that similar effects can be observed in human behavior, which contrasts to long-standing assumptions about human language processing.

In this thesis, I showed that this general property of SRNs can be linked to actual behavioral data in different aspects. In turn, the models revealed a number of properties that are worth further investigation. For example, the results from SRN experiment 2 revealed a rather surprising difference between singular and plural conditions which was not expected. On the one hand, it can be argued that this difference is an artifact of the underlying training grammar, however, although the grammar does not incorporate all possible structures of a language, it only generates possible structures. For example, the fact that the ambiguous forms can only occur as singular forms is not an artificial aspect of the grammar, but mirrors a property in natural language. Thus, it may well be that this difference plays a role with regard to the mental representation and processing of number agreement. A further prediction that can be tested directly is if the plural control conditions used in SRN experiment 2 do indeed inhibit local coherence, as the model suggests.

Finally, the evaluation of number agreement revealed both surprising effects but, as well, interesting cases of garden-paths. Thus, to identifying potentially interesting linguistic structures by evaluating network performance is a nice side-effect of the approach adopted here.

5.3 Self-Consistency, Rational Parsing and SRNs

One of the crucial insights effects of local syntactic coherence provide is that sentence processing is not necessarily self-consistent, neither with regard to syntactic processing, nor with regard to interpretation. However, as discussed in detail in section 3.6.2, the results do not readily answer the question whether human sentence processing is *rational*. In some sense, the approaches that attempt to provide explanations within a rational model thereby imply that a strictly self-consistent parser is "irrational".

With the results reported here, I cannot readily rule out that effects of local syntactic coherence can be explained as results of some rational mechanism. The stance taken here, however, was rather to show that the effects result from the inherent mechanisms of the system, and not to motivate their existence as particularly useful. Whether this kind of system can be described as implementing particular strategies that can be termed as rational, is an interesting question. However, it also shows the problem that the different models trying to account for effects of local coherence vary considerably with regard to their level of description, and therefore also with the particular kind of question they ask.

To conclude, it can be said with some certainty that effects of local syntactic coherence arise, that local coherence can induce interpretation, and that the mechanisms that give rise to the effects are fundamentally based on experience. Whether the effects are mere side-effects of the general properties of the system or rather reflect particular strategies cannot be readily answered. I argued for the first view, assuming that this kind of explanation provides a more parsimonious account and embeds the phenomenon in a general framework of cognition.

6 Summary and Conclusion

Effects of local syntactic coherence have received a fair amount of attention in recent psycholinguistic research, since they seem to question fundamental assumptions about human sentence processing. I started my thesis with three questions:

1. Can the empirical results reported so far be taken as evidence for influence of local syntactic coherence at all?
2. If so, does local syntactic coherence merely pose processing difficulty, or can it even trigger interpretation of the local sequence?
3. Do the effects provide support for a dynamical system perspective on human cognition, as it is, for example, implemented in connectionist approaches of sentence processing?

I argued that local syntactic coherence in sentences can indeed, in contrast to the predictions of a number of current processing models, induce interpretational processes that are in conflict with the global analysis of the sentence they are embedded in. I claimed that this is an effect of frequency, as a result from frequent encounters with the local sequences in contexts where the local interpretation is the correct one. I argued further that the effects can be best explained within an interactive and experience-based framework of human language processing that is not constricted to the self-consistency assumption underlying the majority of past and current models. Approaches that view the human language processing system, and the cognitive system in general, as a dynamical system operating on continuous representations of linguistic information provide a particularly promising framework to incorporate effects of local syntactic coherence without stipulating particular processes or modules. In contrast, the effects are a natural outcome of the underlying mechanisms of acquisition and processing of language.

To support this hypothesis I revisited two visual world experiments providing evidence that local syntactic coherence induces interpretation of the local sequence, reflected in fixations on a scene that is not compatible with the global sentence input.

Strong support for the hypothesis that processing and interpretation of locally coherent sequences is an effect of frequency was provided in a corpus study in section 3.5. The finiteness bias *FB* of the ambiguous elements in the local sequences increases the time spent on the depicted local meaning – but only if local coherence is not inhibited by the preceding words, indicating that the effects result from frequent experience with similar sequences in other contexts.

Chapter 4 provided a cognitive model of the observed results. A simple recurrent network model was developed as an instantiation of a radically experience-based, dynamical model of sentence processing. To evaluate whether the networks are able to learn crucial properties of the training grammar, I developed a method to accurately access performance on long-distance number agreement. In two SRN experiments with test sentences implementing a design derived from visual world experiment 2, effects of local coherence and distributional properties were measured as false-alarm activations of ungrammatical but locally expected words. The results revealed the same pattern that has been observed in the visual world data, but also elicited several effects that may provide interesting starting points for further research on the topic.

To conclude, both the experimental as well as the modeling results provide strong evidence for a dynamical system perspective on human sentence processing, and the crucial influence of distributional properties on both sentence processing and mental representation of linguistic knowledge.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036-1060.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Erlbaum.

Arai, M., & Keller, F. (2012). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67-82.

Baayen, R. H. (2009). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics" [Computer software manual]. (R package version 0.955)

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1181–1182.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. (R package version 0.999999-2)

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (Vol. 279, pp. 279–362). New York: Wiley.

Bicknell, K., & Levy, R. (2009). A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of human language technologies: The 2009 annual conference of the North American*

Chapter of the Association for Computational Linguistics (pp. 665–673). Stroudsburg, PA: Association for Computational Linguistics.

Bicknell, K., Levy, R., & Demberg, V. (2009). Correcting the incorrect: Local coherence effects modeled with prior belief update. In *Proceedings of the 35th Annual Meeting of the Berkeley Linguistics Society* (pp. 13–24). Stroudsburg, PA: Association for Computational Linguistics.

Boston, M. F., Hale, J. T., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1–12.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77-126.

Chen, Z., Jäger, L., & Hale, J. T. (2012). Uncertainty reduction as a predictor of reading difficulty in Chinese relative clauses. In Y.-O. Biq & L. Chen (Eds.), *Proceedings of the 13th International Symposium on Chinese Languages and Linguistics (IsCLL-13)* (pp. 245–261). Taipei, Taiwan: National Taiwan Normal University.

Cho, P. W., Szkudlarek, E., Kukona, A., & Tabor, W. (2011). An artificial grammar investigation into the mental encoding of syntactic structure. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Conference* (pp. 1679–1684). Austin, TX: Cognitive Science Society.

Chomsky, N. (1957). *Syntactic structures*. Den Haag: Mouton.

Chomsky, N. (2002). *On nature and language* (L. Rizzi & A. Belletti, Eds.). Cambridge University Press.

Christiansen, M. H., & Chater, N. (1999a). Connectionist natural language processing: The state of the art. *Cognitive Science: A Multidisciplinary Journal*, *23*(4), 417-437.

Christiansen, M. H., & Chater, N. (1999b). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science: A Multidisciplinary Journal*, *23*(2), 157-205.

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Corley, S., & Crocker, M. W. (2000). The modular statistical hypothesis: Exploring lexical category ambiguity. In M. W. Crocker, M. J. Pickering, & C. Clifton (Eds.), *Architectures and Mechanisms for Language Processing* (pp. 135–160). Cambridge University Press.

Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, *29*, 647-669.

Crocker, M. W., & Corley, S. (2002). Modular architectures and statistical mechanisms: The case from lexical category disambiguation. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing.* Amsterdam: John Benjamins.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Demberg, V., & Keller, F. (2008). Data from eyetracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193-210.

Demberg-Winterfors, V. (2010). *A broad-coverage model of prediction in human sentence processing*. Unpublished doctoral dissertation, University of Edinburgh.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*, 108-127.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science: A Multidisciplinary Journal*, *14*(2), 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71-99.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Fedorenko, E., Piantadosi, S., & Gibson, E. (2012). Processing relative clauses in supportive contexts. *Cognitive Science*, *36*(3), 471–497.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11-15.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1-27.

Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (p. 1139-1144). Austin, TX: Cognitive Science Society.

Frank, S. L. (2010). Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 81–89). Stroudsburg, PA: Association for Computational Linguistics.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading.* (p. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frazier, L., & Clifton Jr., C. (1996). *Construal.* Cambridge: The MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: The MIT Press.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, *54*(3), 363-388.

Gibson, E., & Pearlmutter, N. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, *29*, 231–240.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*(6), 1304-1321.

Gordon, P. C., & Lowder, M. W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Language and Linguistics Compass*, *6*(7), 403–415.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290.

Hahn, L. (2012). Measuring local context as context–word probabilities. *Behavior Research Methods*, *44*(2), 344–360.

Hale, J. T. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on language technologies* (pp. 159–166). Stroudsburg, PA: Association for Computational Linguistics.

Hale, J. T. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101-123.

Hale, J. T. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Hale, J. T. (2011). What a rational parser would do. *Cognitive Science*, *35*(3), 399–443.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, *4*(2), 100–107.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., et al. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10-20.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137–194.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37-55.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*, 365–392.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153-168.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15-47.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, *30*(5), 580-602.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627-645.

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (p. 1178-1183). Mahwah, NJ: Erlbaum.

Konieczny, L., & Doering, P. (2003). Anticipation of Clause-Final Heads. Evidence from Eye-Tracking and SRNs. In *Proceedings of the 4th International Conference on Cognitive Science* (pp. 330–335). Sydney, Australia: University of New South Wales.

Konieczny, L., & Müller, D. (2006). Local coherence interpretation in spoken language: Evidence from a visual world experiment.
(Poster presented at AMLaP 2006: Architectures and Mechanisms of Language Processing. August 30 - September 1, Nijmegen, the Netherlands)

Konieczny, L., & Müller, D. (2010). Das sprachliche Arbeitsgedächtnis: Kapazität oder Erfahrung? *Psychologische Rundschau*, *61*(1), 43-50.

Konieczny, L., Müller, D., Baumann, P., Hachmann, W., & Wolfer, S. A. (2009). Local syntactic coherence interpretation, and how prosody modulates it.
(Poster presented at the 22nd Annual Meeting of the CUNY Conference on Human Sentence Processing. March 26th-28th 2009, Davis, CA)

Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. A. (2009). Local syntactic coherence interpretation. Evidence from a visual world study. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the the 31st Annual Conference of the Cognitive Science Society* (pp. 1133–1138). Austin, TX: Cognitive Science Society.

Konieczny, L., Ruh, N., & Müller, D. (2009). What's in an error? A closer look at SRNs processing relative clauses. In A. Howes, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the the 9th International Conference on Cognitive Modeling.* Manchester, UK.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42.

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*(6), 1009-1051.

Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 234–243). Stroudsburg, PA: Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086-21090.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375-419.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35-54.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676-703.

MacDonald, M. C., Seidenberg, M. S., Traxler, M. J., & Gernsbacher, M. A. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In *Handbook of Psycholinguistics* (2nd ed., pp. 581–611). London: Academic Press.

Marslen-Wilson, W. (1973, 24). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Mayberry III., M. R., Crocker, M. W., & Knoeferle, P. (2005). A connectionist model of sentence comprehension in visual worlds. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1437–1442). Mahwah, NJ: Erlbaum.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, *17*(11), 748-762.

Morgan, E., Keller, F., & Steedman, M. (2010). A bottom-up parsing model of local coherence effects. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1559–1564). Austin, TX: Cognitive Science Society.

Müller, D., Hachmann, W., & Konieczny, L. (2009). Local syntactic and semantic coherence in reading.
(Poster presented at the 15th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP). 7th-9th September 2009, Barcelona, Spain)

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Pickering, M. J., van Gompel, R. P. G., Traxler, M. J., & Gernsbacher, M. A. (2006). Syntactic parsing. In *Handbook of Psycholinguistics* (2nd ed., pp. 455–503). London: Academic Press.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer-Verlag.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191-201.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 324–333). Stroudsburg, PA: Association for Computational Linguistics.

Rodrigruez, P., & Elman, J. (1999). Watching the transients: Viewing a simple recurrent network as a limited counter. *Behaviormetrika*, *26*(1), 51-74.

Rohde, D. L. T. (1999). *The simple language generator: Encoding complex languages with simple grammars.* Pittsburgh, PA.

Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production.* Unpublished doctoral dissertation, Carnegie Mellon University.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition*, *72*(1), 67 - 109.

Ruh, N., Klöckner, K., & Konieczny, L. (2002). Reassessing the ability of simple recurrent networks to account for verbal working memory performance.
(Poster presented at the 15th annual CUNY conference on Human Sentence Processing. March 21st-23th, New York, USA.)

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–30.

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In . T. S. L. Carlson C. Hölscher (Ed.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1637–1642). Austin, TX: Cognitive Science Society.

Spivey, M. J. (2007). *The continuity of mind.* Oxford University Press.

Stabler, E. (1994). The finite connectivity of linguistic structure. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Hillsdale, N.J: Erlbaum.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659.

Tabor, W., Cho, P. W., & Szkudlarek, E. (2012). Fractal unfolding: A metamorphic approach to learning to parse recursive structure. In *Proceedings of the 3rd workshop on cognitive modeling and computational linguistics* (pp. 41–50). Stroudsburg, PA: Association for Computational Linguistics.

Tabor, W., Cho, P. W., & Szkudlarek, E. (2013). Fractal analysis illuminates the form of connectionist structural gradualness. *Topics in Cognitive Science*, *5*(3), 634–667.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355-370.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(2), 431-450.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*, 211-271.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, *23*(4), 491-515.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Trueswell, J., & Tanenhaus, M. K. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (p. 155-179). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407– 430.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284-307.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225–258.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, *82*(4), 767-794.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105-143.

Vosse, T., & Kempen, G. (2009). In defense of competition during syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, *38*(1), 1-9.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: The MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Weldle, H. (2011). *Syntaktische und referenzielle Sprachverstehensprozesse aus konnektionistischer Perspektive*. Phd thesis, University of Freiburg.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250-271.

A List of Figures

B List of Tables

C Stimuli Materials

C.1 Stimuli Visual World Experiment 1

1.1.
Die Tatsache, dass die Frau amüsiert den Säugling der Nachbarin fütterte, war verblüffend.
1.2.
Die Tatsache, dass die Frau lächelnd den Säugling der Nachbarin fütterte, war verblüffend.
1.3.
Die Tatsache, dass die Frau gerade amüsiert den Säugling der Nachbarin fütterte, war verblüffend.
1.4.
Die Tatsache, dass die Frau gerade lächelnd den Säugling der Nachbarin fütterte, war verblüffend.

2.1.
Die Nachricht, dass die Blondine beleidigt den Punker im Stadtgarten schubste, ist überraschend.
2.2.
Die Nachricht, dass die Blondine gekränkt den Punker im Stadtgarten schubste, ist überraschend.
2.3.
Die Nachricht, dass die Blondine gerade beleidigt den Punker im Stadtgarten schubste, ist überraschend.
2.4.
Die Nachricht, dass die Blondine gerade gekränkt den Punker im Stadtgarten schubste, ist überraschend.

3.1.
Die Feststellung, dass der Clown belustigt das Mädchen auf der Straße verjagte, entspricht der Wahrheit.
3.2.
Die Feststellung, dass der Clown fröhlich das Mädchen auf der Straße verjagte, entspricht der Wahrheit.
3.3.
Die Feststellung, dass der Clown gerade belustigt das Mädchen auf der Straße verjagte, entspricht der Wahrheit.
3.4.
Die Feststellung, dass der Clown gerade fröhlich das Mädchen auf der Straße verjagte, entspricht der Wahrheit.

4.1.
Die Nachricht, dass der Trainer ermutigt den Sportler seiner Mannschaft anfeuerte, löste Freude aus.
4.2.

Die Nachricht, dass der Trainer angespornt den Sportler seiner Mannschaft anfeuerte, löste Freude aus.
4.3.
Die Nachricht, dass der Trainer gerade ermutigt den Sportler seiner Mannschaft anfeuerte, löste Freude aus.
4.4.
Die Nachricht, dass der Trainer gerade angespornt den Sportler seiner Mannschaft anfeuerte, löste Freude aus.

5.1.
Die Annahme, dass das Kind beunruhigt die Mutter seiner Freundin anruft, war verständlich.
5.2.
Die Annahme, dass das Kind ängstlich die Mutter seiner Freundin anruft, war verständlich.
5.3.
Die Annahme, dass das Kind gerade beunruhigt die Mutter seiner Freundin anruft, war verständlich.
5.4.
Die Annahme, dass das Kind gerade ängstlich die Mutter seiner Freundin anruft, war verständlich.

6.1.
Die Aussage, dass der Direktor entsetzt die Kassiererin des Supermarkts fixierte, war erstaunlich.
6.2.
Die Aussage, dass der Direktor erschrocken die Kassiererin des Supermarkts fixierte, war erstaunlich.
6.3.
Die Aussage, dass der Direktor gerade entsetzt die Kassiererin des Supermarkts fixierte, war erstaunlich.
6.4.
Die Aussage, dass der Direktor gerade erschrocken die Kassiererin des Supermarkts fixierte, war erstaunlich.

7.1.
Die Auffassung, dass die Radfahrerin entzückt den Jongleur des Zirkus befragte, entsprach den Tatsachen.
7.2.
Die Auffassung, dass die Radfahrerin hingerissen den Jongleur des Zirkus befragte, entsprach den Tatsachen.
7.3.
Die Auffassung, dass die Radfahrerin gerade entzückt den Jongleur des Zirkus befragte, entsprach den Tatsachen.
7.4.
Die Auffassung, dass die Radfahrerin gerade hingerissen den Jongleur des Zirkus befragte, entsprach den Tatsachen.

8.1.
Die Überzeugung, dass der Kapitän erfreut den Musiker des Jahres vorstellte, war interessant.
8.2.
Die Überzeugung, dass der Kapitän freudig den Musiker des Jahres vorstellte, war interessant.
8.3.
Die Überzeugung, dass der Kapitän gerade erfreut den Musiker des Jahres vorstellte, war interessant.
8.4.

Die Überzeugung, dass der Kapitän gerade freudig den Musiker des Jahres vorstellte, war interessant.

9.1.
Die Mitteilung, dass der Einbrecher erschreckt das Kind der Hausbewohner anstarrte, sorgte für Erheiterung.
9.2.
Die Mitteilung, dass der Einbrecher fassungslos das Kind der Hausbewohner anstarrte, sorgte für Erheiterung.
9.3.
Die Mitteilung, dass der Einbrecher gerade erschreckt das Kind der Hausbewohner anstarrte, sorgte für Erheiterung.
9.4.
Die Mitteilung, dass der Einbrecher gerade fassungslos das Kind der Hausbewohner anstarrte, sorgte für Erheiterung.

10.1.
Die Beobachtung, dass der Feuerwehrmann fasziniert die Polizistin des Reviers betrachtete, war offenkundig.
10.2.
Die Beobachtung, dass der Feuerwehrmann verliebt die Polizistin des Reviers betrachtete, war offenkundig.
10.3.
Die Beobachtung, dass der Feuerwehrmann gerade fasziniert die Polizistin des Reviers betrachtete, war offenkundig.
10.4.
Die Beobachtung, dass der Feuerwehrmann gerade verliebt die Polizistin des Reviers betrachtete, war offenkundig.

11.1.
Die Tatsache, dass der Briefträger verängstigt den Wachhund des Hauses streichelte, war offensichtlich.
11.2.
Die Tatsache, dass der Briefträger ängstlich den Wachhund des Hauses streichelte, war offensichtlich.
11.3.
Die Tatsache, dass der Briefträger gerade verängstigt den Wachhund des Hauses streichelte, war offensichtlich.
11.4.
Die Tatsache, dass der Briefträger gerade ängstlich den Wachhund des Hauses streichelte, war offensichtlich.

12.1.
Die Nachricht, dass der Detektiv erzürnt den Kunden des Kaufhauses ertappte, beruhte auf Tatsachen.
12.2.
Die Nachricht, dass der Detektiv zornig den Kunden des Kaufhauses ertappte, beruhte auf Tatsachen.
12.3.
Die Nachricht, dass der Detektiv gerade erzürnt den Kunden des Kaufhauses ertappte, beruhte auf Tatsachen.
12.4.
Die Nachricht, dass der Detektiv gerade zornig den Kunden des Kaufhauses ertappte, beruhte auf Tatsachen.

13.1.
Die Mitteilung, dass die Ärztin frustriert den Kranken in seinem Zimmer untersuchte, machte die Runde.
13.2.
Die Mitteilung, dass die Ärztin gefrustet den Kranken in seinem Zimmer untersuchte, machte die Runde.
13.3.
Die Mitteilung, dass die Ärztin gerade frustriert den Kranken in seinem Zimmer untersuchte, machte die Runde.
13.4.
Die Mitteilung, dass die Ärztin gerade gefrustet den Kranken in seinem Zimmer untersuchte, machte die Runde.

14.1.
Die Information, dass der Weihnachtmann verwundert den Schornsteinfeger auf dem Dach begrüßte, war interessant.
14.2.
Die Information, dass der Weihnachtmann fassungslos den Schornsteinfeger auf dem Dach begrüßte, war interessant.
14.3.
Die Information, dass der Weihnachtmann gerade verwundert den Schornsteinfeger auf dem Dach begrüßte, war interessant.
14.4.
Die Information, dass der Weihnachtmann gerade fassungslos den Schornsteinfeger auf dem Dach begrüßte, war interessant.

15.1.
Die Beobachtung, dass der Schüler entmutigt den Lehrer der Klasse anschaute, war interessant.
15.2.
Die Beobachtung, dass der Schüler mutlos den Lehrer der Klasse anschaute, war interessant.
15.3.
Die Beobachtung, dass der Schüler gerade entmutigt den Lehrer der Klasse anschaute, war interessant.
15.4.
Die Beobachtung, dass der Schüler gerade mutlos den Lehrer der Klasse anschaute, war interessant.

16.1.
Die Vermutung, dass der Opa verdutzt die Enkelin auf der Familienfeier begrüßte, entsprach der Wahrheit.
16.2.
Die Vermutung, dass der Opa verdattert die Enkelin auf der Familienfeier begrüßte, entsprach der Wahrheit.
16.3.
Die Vermutung, dass der Opa gerade verdutzt die Enkelin auf der Familienfeier begrüßte, entsprach der Wahrheit.
16.4.
Die Vermutung, dass der Opa gerade verdattert die Enkelin auf der Familienfeier begrüßte, entsprach der Wahrheit.

17.1.

Die Feststellung, dass die Floristin überrascht den Reporter des Radiosenders entdeckte, bestätigte sich.
17.2.
Die Feststellung, dass die Floristin perplex den Reporter des Radiosenders entdeckte, bestätigte sich.
17.3.
Die Feststellung, dass die Floristin gerade überrascht den Reporter des Radiosenders entdeckte, bestätigte sich.
17.4.
Die Feststellung, dass die Floristin gerade perplex den Reporter des Radiosenders entdeckte, bestätigte sich.

18.1.
Die Annahme, dass der alte Mann verzaubert die Kellnerin des Restaurants anstarrte, war nahe liegend.
18.2.
Die Annahme, dass der alte Mann träumerisch die Kellnerin des Restaurants anstarrte, war nahe liegend.
18.3.
Die Annahme, dass der alte Mann gerade verzaubert die Kellnerin des Restaurants anstarrte, war nahe liegend.
18.4.
Die Annahme, dass der alte Mann gerade träumerisch die Kellnerin des Restaurants anstarrte, war nahe liegend.

19.1.
Die Tatsache, dass der Moderator verschüchtert die Sängerin des Ensembles befragte, war offenkundig.
19.2.
Die Tatsache, dass der Moderator schüchtern die Sängerin des Ensembles befragte, war offenkundig.
19.3.
Die Tatsache, dass der Moderator gerade verschüchtert die Sängerin des Ensembles befragte, war offenkundig.
19.4.
Die Tatsache, dass der Moderator gerade schüchtern die Sängerin des Ensembles befragte, war offenkundig.

20.1.
Die Beobachtung, dass der Autofahrer irritiert das Model aus Paris anblickte, überraschte nicht.
20.2.
Die Beobachtung, dass der Autofahrer staunend das Model aus Paris anblickte, überraschte nicht.
20.3.
Die Beobachtung, dass der Autofahrer gerade irritiert das Model aus Paris anblickte, überraschte nicht.
20.4.
Die Beobachtung, dass der Autofahrer gerade staunend das Model aus Paris anblickte, überraschte nicht.

21.1.
Die Behauptung, dass der Junge verschmutzt die Katze des Nachbarn streichelte, wurde bestätigt.
21.2.
Die Behauptung, dass der Junge schmutzig die Katze des Nachbarn streichelte, wurde bestätigt.
21.3.
Die Behauptung, dass der Junge gerade verschmutzt die Katze des Nachbarn streichelte, wurde bestätigt.

21.4.
Die Behauptung, dass der Junge gerade schmutzig die Katze des Nachbarn streichelte, wurde bestätigt.

22.1.
Die Feststellung, dass die Astronautin erstaunt den Außerirdischen vom Mars aufstöberte, entsprach der Wahrheit.
22.2.
Die Feststellung, dass die Astronautin ungläubig den Außerirdischen vom Mars aufstöberte, entsprach der Wahrheit.
22.3.
Die Feststellung, dass die Astronautin gerade erstaunt den Außerirdischen vom Mars aufstöberte, entsprach der Wahrheit.
22.4.
Die Feststellung, dass die Astronautin gerade ungläubig den Außerirdischen vom Mars aufstöberte, entsprach der Wahrheit.

23.1.
Die Bemerkung, dass der Schaffner verärgert den Fahrgast des Schnellzuges hinauswarf, sorgte für Heiterkeit.
23.2.
Die Bemerkung, dass der Schaffner ärgerlich den Fahrgast des Schnellzuges hinauswarf, sorgte für Heiterkeit.
23.3.
Die Bemerkung, dass der Schaffner gerade verärgert den Fahrgast des Schnellzuges hinauswarf, sorgte für Heiterkeit.
23.4.
Die Bemerkung, dass der Schaffner gerade ärgerlich den Fahrgast des Schnellzuges hinauswarf, sorgte für Heiterkeit.

24.1.
Die Behauptung, dass der Straßenfeger erheitert den Wanderer aus der Eifel musterte, lieferte Grund für Diskussionen.
24.2.
Die Behauptung, dass der Straßenfeger lachend den Wanderer aus der Eifel musterte, lieferte Grund für Diskussionen.
24.3.
Die Behauptung, dass der Straßenfeger gerade erheitert den Wanderer aus der Eifel musterte, lieferte Grund für Diskussionen.
24.4.
Die Behauptung, dass der Straßenfeger gerade lachend den Wanderer aus der Eifel musterte, lieferte Grund für Diskussionen.

25.1.
Die Mutmaßung, dass die Krankenschwester verblüfft den Patienten der Intensivstation verband, bewahrheitete sich.

25.2.
Die Mutmaßung, dass die Krankenschwester sprachlos den Patienten der Intensivstation verband, bewahrheitete sich.
25.3.
Die Mutmaßung, dass die Krankenschwester gerade verblüfft den Patienten der Intensivstation verband, bewahrheitete sich.
25.4.
Die Mutmaßung, dass die Krankenschwester gerade sprachlos den Patienten der Intensivstation verband, bewahrheitete sich.

26.1.
Die Information, dass der Gärtner betört die Gräfin des Landgutes angaffte, war aufschlussreich.
26.2.
Die Information, dass der Gärtner schmachtend die Gräfin des Landgutes angaffte, war aufschlussreich.
26.3.
Die Information, dass der Gärtner gerade betört die Gräfin des Landgutes angaffte, war aufschlussreich.
26.4.
Die Information, dass der Gärtner gerade schmachtend die Gräfin des Landgutes angaffte, war aufschlussreich.

27.1.
Der Bericht, dass der Großvater begeistert die Braut seines Sohnes küsste, sorgte für Aufruhr.
27.2.
Der Bericht, dass der Großvater glücklich die Braut seines Sohnes küsste, sorgte für Aufruhr.
27.3.
Der Bericht, dass der Großvater gerade begeistert die Braut seines Sohnes küsste, sorgte für Aufruhr.
27.4.
Der Bericht, dass der Großvater gerade glücklich die Braut seines Sohnes küsste, sorgte für Aufruhr.

28.1.
Die Prophezeiung, dass der Fakir hypnotisiert die Schlange im Korb anstierte, beruhte auf Tatsachen.
28.2.
Die Prophezeiung, dass der Fakir gebannt die Schlange im Korb anstierte, beruhte auf Tatsachen.
28.3.
Die Prophezeiung, dass der Fakir gerade hypnotisiert die Schlange im Korb anstierte, beruhte auf Tatsachen.
28.4.
Die Prophezeiung, dass der Fakir gerade gebannt die Schlange im Korb anstierte, beruhte auf Tatsachen.

29.1.
Die Annahme, dass der Bergmann verdreckt das Baby seiner Nichte hochhob, sorgte für Missstimmung.
29.2.
Die Annahme, dass der Bergmann dreckig das Baby seiner Nichte hochhob, sorgte für Missstimmung.

29.3.
Die Annahme, dass der Bergmann gerade verdreckt das Baby seiner Nichte hochhob, sorgte für Missstimmung.
29.4.
Die Annahme, dass der Bergmann gerade dreckig das Baby seiner Nichte hochhob, sorgte für Missstimmung.

30.1.
Die Beobachtung, dass der Polizist erniedrigt die Demonstrierende der Studierendenbewegung ansah, war erstaunlich.
30.2.
Die Beobachtung, dass der Polizist gedemütigt die Demonstrierende der Studierendenbewegung ansah, war erstaunlich.
30.3.
Die Beobachtung, dass der Polizist gerade erniedrigt die Demonstrierende der Studierendenbewegung ansah, war erstaunlich.
30.4.
Die Beobachtung, dass der Polizist gerade gedemütigt die Demonstrierende der Studierendenbewegung ansah, war erstaunlich.

31.1.
Die Feststellung, dass der Bauer schockiert die Magd des Nachbarhofs erkannte, entsprach der Wahrheit.
31.2.
Die Feststellung, dass der Bauer geschockt die Magd des Nachbarhofs erkannte, entsprach der Wahrheit.
31.3.
Die Feststellung, dass der Bauer gerade schockiert die Magd des Nachbarhofs erkannte, entsprach der Wahrheit.
31.4.
Die Feststellung, dass der Bauer gerade geschockt die Magd des Nachbarhofs erkannte, entsprach der Wahrheit.

32.1.
Die Vermutung, dass die Chefin enttäuscht den Angestellten ihrer Firma verabschiedete, bestätigte sich.
32.2.
Die Vermutung, dass die Chefin traurig den Angestellten ihrer Firma verabschiedete, bestätigte sich.
32.3.
Die Vermutung, dass die Chefin gerade enttäuscht den Angestellten ihrer Firma verabschiedete, bestätigte sich.
32.4.
Die Vermutung, dass die Chefin gerade traurig den Angestellten ihrer Firma verabschiedete, bestätigte sich.

33.1.
Der Sachverhalt, dass die Stewardess verstört den Piloten des Flugunternehmens bediente, ist nachvollziehbar.
33.2.
Der Sachverhalt, dass die Stewardess betreten den Piloten des Flugunternehmens bediente, ist nachvollziehbar.
33.3.

Der Sachverhalt, dass die Stewardess gerade verstört den Piloten des Flugunternehmens bediente, ist nachvollziehbar.
33.4.
Der Sachverhalt, dass die Stewardess gerade betreten den Piloten des Flugunternehmens bediente, ist nachvollziehbar.

34.1.
Die Nachricht, dass das Mädchen zerkratzt den Bruder ihrer Freundin fragte, löste Verwunderung aus.
34.2.
Die Nachricht, dass das Mädchen geschunden den Bruder ihrer Freundin fragte, löste Verwunderung aus.
34.3.
Die Nachricht, dass das Mädchen gerade zerkratzt den Bruder ihrer Freundin fragte, löste Verwunderung aus.
34.4.
Die Nachricht, dass das Mädchen gerade geschunden den Bruder ihrer Freundin fragte, löste Verwunderung aus.

35.1.
Die Beobachtung, dass die Tänzerin erregt den Gast des Lokals aufforderte, war interessant.
35.2.
Die Beobachtung, dass die Tänzerin aufgeregt den Gast des Lokals aufforderte, war interessant.
35.3.
Die Beobachtung, dass die Tänzerin gerade erregt den Gast des Lokals aufforderte, war interessant.
35.4.
Die Beobachtung, dass die Tänzerin gerade aufgeregt den Gast des Lokals aufforderte, war interessant.

36.1.
Die Nachricht, dass der Bankräuber vermummt die Angestellte des Kreditinstitutes bedrohte, löste Entsetzen aus.
36.2.
Die Nachricht, dass der Bankräuber getarnt die Angestellte des Kreditinstitutes bedrohte, löste Entsetzen aus.
36.3.
Die Nachricht, dass der Bankräuber gerade vermummt die Angestellte des Kreditinstitutes bedrohte, löste Entsetzen aus.
36.4.
Die Nachricht, dass der Bankräuber gerade getarnt die Angestellte des Kreditinstitutes bedrohte, löste Entsetzen aus.

37.1.
Die Legende, dass der Ritter verwundet den Drachen des Berges bekämpfte, erfreute sich großer Beliebtheit.
37.2.
Die Legende, dass der Ritter blutend den Drachen des Berges bekämpfte, erfreute sich großer Beliebtheit.
37.3.
Die Legende, dass der Ritter gerade verwundet den Drachen des Berges bekämpfte, erfreute sich großer Be-

liebtheit.
37.4.
Die Legende, dass der Ritter gerade blutend den Drachen des Berges bekämpfte, erfreute sich großer Beliebtheit.

38.1.
Die Feststellung, dass der Fußballer verletzt den Schiedsrichter des Turniers rief, erstaunte die Zuschauer.
38.2.
Die Feststellung, dass der Fußballer angeschlagen den Schiedsrichter des Turniers rief, erstaunte die Zuschauer.
38.3.
Die Feststellung, dass der Fußballer gerade verletzt den Schiedsrichter des Turniers rief, erstaunte die Zuschauer.
38.4.
Die Feststellung, dass der Fußballer gerade angeschlagen den Schiedsrichter des Turniers rief, erstaunte die Zuschauer.

39.1.
Die Tatsache, dass die Tante verkleidet das Kind ihrer Schwester begleitete, war beruhigend.
39.2.
Die Tatsache, dass die Tante als Fee das Kind ihrer Schwester begleitete, war beruhigend.
39.3.
Die Tatsache, dass die Tante verkleidet das Kind ihrer Schwester begleitete, war beruhigend.
39.4.
Die Tatsache, dass die Tante als Fee das Kind ihrer Schwester begleitete, war beruhigend.

40.1.
Die Feststellung, dass der Künstler bekleckst die Tochter des Direktors porträtierte, sorgte für Empörung.
40.2.
Die Feststellung, dass der Künstler angemalt die Tochter des Direktors porträtierte, sorgte für Empörung.
40.3.
Die Feststellung, dass der Künstler gerade bekleckst die Tochter des Direktors porträtierte, sorgte für Empörung.
40.4.
Die Feststellung, dass der Künstler gerade angemalt die Tochter des Direktors porträtierte, sorgte für Empörung.

41.1.
Die Nachricht, dass die Schwester bespritzt den Bruder im Anzug umarmt, ist amüsant.
41.2.
Die Nachricht, dass die Schwester dreckig den Bruder im Anzug umarmt, ist amüsant.
41.3.
Die Nachricht, dass die Schwester gerade bespritzt den Bruder im Anzug umarmt, ist amüsant.
41.4.
Die Nachricht, dass die Schwester gerade dreckig den Bruder im Anzug umarmt, ist amüsant.

42.1.
Die Beobachtung, dass der Physiotherapeut bewegt die Hand seiner Patientin küsst, ist überraschend.
42.2.
Die Beobachtung, dass der Physiotherapeut gerührt die Hand seiner Patientin küsst, ist überraschend.
42.3.
Die Beobachtung, dass der Physiotherapeut gerade bewegt die Hand seiner Patientin küsst, ist überraschend.
42.4.
Die Beobachtung, dass der Physiotherapeut gerade gerührt die Hand seiner Patientin küsst, ist überraschend.

43.1.
Die Mutmaßung, dass der Maurermeister erbost den Architekten der Wohnanlage herbeiwinkt, ist wahr.
43.2.
Die Mutmaßung, dass der Maurermeister aufgebracht den Architekten der Wohnanlage herbeiwinkt, ist wahr.
43.3.
Die Mutmaßung, dass der Maurermeister gerade erbost den Architekten der Wohnanlage herbeiwinkt, ist wahr.
43.4.
Die Mutmaßung, dass der Maurermeister gerade aufgebracht den Architekten der Wohnanlage herbeiwinkt, ist wahr.

44.1.
Die Tatsache, dass der Ehemann rasiert die Beine seiner Frau bewundert, ist erfreulich.
44.2.
Die Tatsache, dass der Ehemann bartlos die Beine seiner Frau bewundert, ist erfreulich.
44.3.
Die Tatsache, dass der Ehemann gerade rasiert die Beine seiner Frau bewundert, ist erfreulich.
44.4.
Die Tatsache, dass der Ehemann gerade bartlos die Beine seiner Frau bewundert, ist erfreulich.

45.1.
Die Vermutung, dass der Schüler verpennt den Bus zur Schule erreicht, ist plausibel.
45.2.
Die Vermutung, dass der Schüler müde den Bus zur Schule erreicht, ist plausibel.
45.3.
Die Vermutung, dass der Schüler gerade verpennt den Bus zur Schule erreicht, ist plausibel.
45.4.
Die Vermutung, dass der Schüler gerade müde den Bus zur Schule erreicht, ist plausibel.

46.1.
Die Nachricht, dass der Professor ermüdet den Studierenden im Seminar aufruft, ist überraschend.
46.2.
Die Nachricht, dass der Professor schläfrig den Studierenden im Seminar aufruft, ist überraschend.

46.3.
Die Nachricht, dass der Professor gerade ermüdet den Studierenden im Seminar aufruft, ist überraschend.
46.4.
Die Nachricht, dass der Professor gerade schläfrig den Studierenden im Seminar aufruft, ist überraschend.

47.1.
Die Feststellung, dass die Sportlehrerin erschöpft den Schüler ihrer Klasse ermahnt, ist offenkundig.
47.2.
Die Feststellung, dass die Sportlehrerin matt den Schüler ihrer Klasse ermahnt, ist offenkundig.
47.3.
Die Feststellung, dass die Sportlehrerin gerade erschöpft den Schüler ihrer Klasse ermahnt, ist offenkundig.
47.4.
Die Feststellung, dass die Sportlehrerin gerade matt den Schüler ihrer Klasse ermahnt, ist offenkundig.

48.1.
Die Beobachtung, dass der Pastor überzeugt den Bettler an der Tür abweist, ist erstaunlich.
48.2.
Die Beobachtung, dass der Pastor entschlossen den Bettler an der Tür abweist, ist erstaunlich.
48.3.
Die Beobachtung, dass der Pastor gerade überzeugt den Bettler an der Tür abweist, ist erstaunlich.
48.4.
Die Beobachtung, dass der Pastor gerade entschlossen den Bettler an der Tür abweist, ist erstaunlich.

C.2 Stimuli Visual World Experiment 2

1.1.
Dieter, dem die Schwimmerin amüsiert den Surfer am Strand präsentiert, freut sich über die Bekanntschaft.
1.2.
Dieter, dem die Schwimmerin heiter den Surfer am Strand präsentiert, freut sich über die Bekanntschaft.
1.3.
Dieter, dem die Schwimmerin äußerst amüsiert den Surfer am Strand präsentiert, freut sich über die Bekanntschaft.
1.4.
Dieter, dem die Schwimmerin äußerst heiter den Surfer am Strand präsentiert, freut sich über die Bekanntschaft.

Item 2 was excluded from analysis due to low accuracy in the clicking task
2.1.
Matthias, dem die Krankenschwester verschüchtert den Besucher aus Dortmund bringt, wartet schon lange.
2.2.
Matthias, dem die Krankenschwester schüchtern den Besucher aus Dortmund bringt, wartet schon lange.
2.3.
Matthias, dem die Krankenschwester äußerst verschüchtert den Besucher aus Dortmund bringt, wartet schon lange.
2.4.
Matthias, dem die Krankenschwester äußerst schüchtern den Besucher aus Dortmund bringt, wartet schon lange.

3.1.
Rolf, dem die Sekretärin frustriert den Anwalt aus Amerika verspricht, hat noch nichts gegessen.
3.2.
Rolf, dem die Sekretärin gefrustet den Anwalt aus Amerika verspricht, hat noch nichts gegessen.
3.3.
Rolf, dem die Sekretärin äußerst frustriert den Anwalt aus Amerika verspricht, hat noch nichts gegessen.
3.4.
Rolf, dem die Sekretärin äußerst gefrustet den Anwalt aus Amerika verspricht, hat noch nichts gegessen.

4.1.
Franz, dem die Ballerina verzaubert den Chef aus Paris vorführt, ist sehr beeindruckt.
4.2.
Franz, dem die Ballerina träumerisch den Chef aus Paris vorführt, ist sehr beeindruckt.
4.3.
Franz, dem die Ballerina äußerst verzaubert den Chef aus Paris vorführt, ist sehr beeindruckt.
4.4.
Franz, dem die Ballerina äußerst träumerisch den Chef aus Paris vorführt, ist sehr beeindruckt.

5.1.
Harald, dem die Magd verärgert den Bettler im Gästezimmer verschweigt, wundert sich über den Mantel.
5.2.
Harald, dem die Magd ärgerlich den Bettler im Gästezimmer verschweigt, wundert sich über den Mantel.
5.3.
Harald, dem die Magd äußerst verärgert den Bettler im Gästezimmer verschweigt, wundert sich über den Mantel.
5.4.
Harald, dem die Magd äußerst ärgerlich den Bettler im Gästezimmer verschweigt, wundert sich über den Mantel.

6.1.
Tobias, dem der Holzfäller erbost den Lehrling aus Augsburg zuteilt, ist gerade erst angekommen.
6.2.
Tobias, dem der Holzfäller aufgebracht den Lehrling aus Augsburg zuteilt, ist gerade erst angekommen.
6.3.
Tobias, dem der Holzfäller äußerst erbost den Lehrling aus Augsburg zuteilt, ist gerade erst angekommen.
6.4.
Tobias, dem der Holzfäller äußerst aufgebracht den Lehrling aus Augsburg zuteilt, ist gerade erst angekommen.

7.1.
Ullrich, dem der Barkeeper belustigt die Tänzerin aus Ulm ankündigt, freut sich auf den Abend.
7.2.
Ullrich, dem der Barkeeper fröhlich die Tänzerin aus Ulm ankündigt, freut sich auf den Abend.
7.3.
Ullrich, dem der Barkeeper äußerst belustigt die Tänzerin aus Ulm ankündigt, freut sich auf den Abend.
7.4.
Ullrich, dem der Barkeeper äußerst fröhlich die Tänzerin aus Ulm ankündigt, freut sich auf den Abend.

8.1.
Elmar, dem die Witwe begeistert den Pudel aus Frankreich vermacht, ist sehr zufrieden.
8.2.
Elmar, dem die Witwe glücklich den Pudel aus Frankreich vermacht, ist sehr zufrieden.
8.3.
Elmar, dem die Witwe äußerst begeistert den Pudel aus Frankreich vermacht, ist sehr zufrieden.
8.4.
Elmar, dem die Witwe äußerst glücklich den Pudel aus Frankreich vermacht, ist sehr zufrieden.

9.1.
Felix, dem die Tante verängstigt den Papagei aus Java vererbt, freut sich sehr.
9.2.
Felix, dem die Tante ängstlich den Papagei aus Java vererbt, freut sich sehr.
9.3.

Felix, dem die Tante äußerst verängstigt den Papagei aus Java vererbt, freut sich sehr.
9.4.
Felix, dem die Tante äußerst ängstlich den Papagei aus Java vererbt, freut sich sehr.

10.1.
Theo, dem die Taucherin überrascht den Tintenfisch aus dem Wasser hinhält, ekelt sich ein wenig.
10.2.
Theo, dem die Taucherin perplex den Tintenfisch aus dem Wasser hinhält, ekelt sich ein wenig.
10.3.
Theo, dem die Taucherin äußerst überrascht den Tintenfisch aus dem Wasser hinhält, ekelt sich ein wenig.
10.4.
Theo, dem die Taucherin äußerst perplex den Tintenfisch aus dem Wasser hinhält, ekelt sich ein wenig.

11.1.
Pascal, dem der Reiter verschmutzt die Zuchtstute aus Brandenburg präsentiert, züchtet schon lange. Pferde.
11.2.
Pascal, dem der Reiter schmutzig die Zuchtstute aus Brandenburg präsentiert, züchtet schon lange. Pferde.
11.3.
Pascal, dem der Reiter äußerst verschmutzt die Zuchtstute aus Brandenburg präsentiert, züchtet schon lange. Pferde.
11.4.
Pascal, dem der Reiter äußerst schmutzig die Zuchtstute aus Brandenburg präsentiert, züchtet schon lange. Pferde.

12.1.
Sascha, dem die Dozentin erschöpft den Praktikanten aus England verspricht, kann es noch nicht glauben.
12.2.
Sascha, dem die Dozentin matt den Praktikanten aus England verspricht, kann es noch nicht glauben.
12.3.
Sascha, dem die Dozentin äußerst erschöpft den Praktikanten aus England verspricht, kann es noch nicht glauben.
12.4.
Sascha, dem die Dozentin äußerst matt den Praktikanten aus England verspricht, kann es noch nicht glauben.

13.1.
Guido, dem der Fakir hypnotisiert die Schlange aus Indien vorführt, schaut gebannt zu.
13.2.
Guido, dem der Fakir gebannt die Schlange aus Indien vorführt, schaut gebannt zu.
13.3.
Guido, dem der Fakir äußerst hypnotisiert die Schlange aus Indien vorführt, schaut gebannt zu.
13.4.
Guido, dem der Fakir äußerst gebannt die Schlange aus Indien vorführt, schaut gebannt zu.

14.1.
Heinrich, dem der Greis überzeugt den Besucher aus Las Vegas weissagt, glaubt nicht so recht daran.
14.2.
Heinrich, dem der Greis entschlossen den Besucher aus Las Vegas weissagt, glaubt nicht so recht daran.
14.3.
Heinrich, dem der Greis äußerst überzeugt den Besucher aus Las Vegas weissagt, glaubt nicht so recht daran.
14.4.
Heinrich, dem der Greis äußerst entschlossen den Besucher aus Las Vegas weissagt, glaubt nicht so recht daran.

15.1.
Ludwig, dem die Polizistin verwundert den Dieb am Bahnhof meldet, ist der Chef der Behörde.
15.2.
Ludwig, dem die Polizistin fassungslos den Dieb am Bahnhof meldet, ist der Chef der Behörde.
15.3.
Ludwig, dem die Polizistin äußerst verwundert den Dieb am Bahnhof meldet, ist der Chef der Behörde.
15.4.
Ludwig, dem die Polizistin äußerst fassungslos den Dieb am Bahnhof meldet, ist der Chef der Behörde.

16.1.
Luis, dem die Bäuerin schockiert den Jungen aus dem Nachbardorf abgibt, ist ein guter Babysitter.
16.2.
Luis, dem die Bäuerin geschockt den Jungen aus dem Nachbardorf abgibt, ist ein guter Babysitter.
16.3.
Luis, dem die Bäuerin äußerst schockiert den Jungen aus dem Nachbardorf abgibt, ist ein guter Babysitter.
16.4.
Luis, dem die Bäuerin äußerst geschockt den Jungen aus dem Nachbardorf abgibt, ist ein guter Babysitter.

17.1.
Leonhard, dem die Schauspielerin beleidigt den Stylisten aus London vorschlägt, ist nicht ganz überzeugt davon.
17.2.
Leonhard, dem die Schauspielerin gekränkt den Stylisten aus London vorschlägt, ist nicht ganz überzeugt davon.
17.3.
Leonhard, dem die Schauspielerin äußerst beleidigt den Stylisten aus London vorschlägt, ist nicht ganz überzeugt davon.
17.4.
Leonhard, dem die Schauspielerin äußerst gekränkt den Stylisten aus London vorschlägt, ist nicht ganz überzeugt davon.

18.1.
Julian, dem die Direktorin verärgert den Nachhilfelehrer in Neustadt aufbrummt, ist 10 Jahre alt.
18.2.

Julian, dem die Direktorin ärgerlich den Nachhilfelehrer in Neustadt aufbrummt, ist 10 Jahre alt.
18.3.
Julian, dem die Direktorin äußerst verärgert den Nachhilfelehrer in Neustadt aufbrummt, ist 10 Jahre alt.
18.4.
Julian, dem die Direktorin äußerst ärgerlich den Nachhilfelehrer in Neustadt aufbrummt, ist 10 Jahre alt.

19.1.
Nils, dem die Trainerin erbost den Masseur im Nachbarort aufzwingt, hat keine Schmerzen mehr.
19.2.
Nils, dem die Trainerin aufgebracht den Masseur im Nachbarort aufzwingt, hat keine Schmerzen mehr.
19.3.
Nils, dem die Trainerin äußerst erbost den Masseur im Nachbarort aufzwingt, hat keine Schmerzen mehr.
19.4.
Nils, dem die Trainerin äußerst aufgebracht den Masseur im Nachbarort aufzwingt, hat keine Schmerzen mehr.

20.1.
Simon, dem die Großmutter begeistert den Hirtenhund aus Schottland kauft, hat morgen Geburtstag.
20.2.
Simon, dem die Großmutter glücklich den Hirtenhund aus Schottland kauft, hat morgen Geburtstag.
20.3.
Simon, dem die Großmutter äußerst begeistert den Hirtenhund aus Schottland kauft, hat morgen Geburtstag.
20.4.
Simon, dem die Großmutter äußerst glücklich den Hirtenhund aus Schottland kauft, hat morgen Geburtstag.

21.1.
Fritz, dem die Artistin beglückt den Clown aus Ungarn vorspielt, schaut begeistert zu.
21.2.
Fritz, dem die Artistin glücklich den Clown aus Ungarn vorspielt, schaut begeistert zu.
21.3.
Fritz, dem die Artistin äußerst beglückt den Clown aus Ungarn vorspielt, schaut begeistert zu.
21.4.
Fritz, dem die Artistin äußerst glücklich den Clown aus Ungarn vorspielt, schaut begeistert zu.

22.1.
Herbert, dem der Förster verletzt den Hirsch aus dem Schwarzwald serviert, hat großen Hunger.
22.2.
Herbert, dem der Förster angeschlagen den Hirsch aus dem Schwarzwald serviert, hat großen Hunger.
22.3.
Herbert, dem der Förster äußerst verletzt den Hirsch aus dem Schwarzwald serviert, hat großen Hunger.
22.4.
Herbert, dem der Förster äußerst angeschlagen den Hirsch aus dem Schwarzwald serviert, hat großen Hunger.

23.1.
Hugo, dem die Astronautin überrascht den Außerirdischen vom Mars zeigt, hat so etwas noch nie gesehen.
23.2.
Hugo, dem die Astronautin ungläubig den Außerirdischen vom Mars zeigt, hat so etwas noch nie gesehen.
23.3.
Hugo, dem die Astronautin äußerst überrascht den Außerirdischen vom Mars zeigt, hat so etwas noch nie gesehen.
23.4.
Hugo, dem die Astronautin äußerst ungläubig den Außerirdischen vom Mars zeigt, hat so etwas noch nie gesehen.

24.1.
Frieder, dem der Verkäufer erfreut den Hund aus der Zoohandlung überreicht, hätte lieber einen Vogel bekommen.
24.2.
Frieder, dem der Verkäufer freudig den Hund aus der Zoohandlung überreicht, hätte lieber einen Vogel bekommen.
24.3.
Frieder, dem der Verkäufer äußerst erfreut den Hund aus der Zoohandlung überreicht, hätte lieber einen Vogel bekommen.
24.4.
Frieder, dem der Verkäufer äußerst freudig den Hund aus der Zoohandlung überreicht, hätte lieber einen Vogel bekommen.

25.1.
Peter, dem die Nachbarin erheitert den Jungen aus dem Dorf mitbringt, freut sich über die Ablenkung.
25.2.
Peter, dem die Nachbarin froh den Jungen aus dem Dorf mitbringt, freut sich über die Ablenkung.
25.3.
Peter, dem die Nachbarin äußerst erheitert den Jungen aus dem Dorf mitbringt, freut sich über die Ablenkung.
25.4.
Peter, dem die Nachbarin äußerst froh den Jungen aus dem Dorf mitbringt, freut sich über die Ablenkung.

26.1.
Martin, dem die Reiseleiterin belustigt den Bäcker im Dorf zeigt, merkt sich den Ort für später.
26.2.
Martin, dem die Reiseleiterin fröhlich den Bäcker im Dorf zeigt, merkt sich den Ort für später.
26.3.
Martin, dem die Reiseleiterin äußerst belustigt den Bäcker im Dorf zeigt, merkt sich den Ort für später.
26.4.
Martin, dem die Reiseleiterin äußerst fröhlich den Bäcker im Dorf zeigt, merkt sich den Ort für später.

27.1.
Anton, dem die Schwester begeistert den Papagei aus Spanien spendiert, redet sonst selten mit Vögeln.
27.2.
Anton, dem die Schwester glücklich den Papagei aus Spanien spendiert, redet sonst selten mit Vögeln.
27.3.
Anton, dem die Schwester äußerst begeistert den Papagei aus Spanien spendiert, redet sonst selten mit Vögeln.
27.4.
Anton, dem die Schwester äußerst glücklich den Papagei aus Spanien spendiert, redet sonst selten mit Vögeln.

28.1.
Jürgen, dem die Lehrerin ermüdet den Tischnachbarn in der Schule zuweist, ist in der dritten Klasse.
28.2.
Jürgen, dem die Lehrerin schläfrig den Tischnachbarn in der Schule zuweist, ist in der dritten Klasse.
28.3.
Jürgen, dem die Lehrerin äußerst ermüdet den Tischnachbarn in der Schule zuweist, ist in der dritten Klasse.
28.4.
Jürgen, dem die Lehrerin äußerst schläfrig den Tischnachbarn in der Schule zuweist, ist in der dritten Klasse.

29.1.
Roland, dem die Pflegerin beunruhigt den Arzt aus der Klinik ruft, ist von der Leiter gefallen.
29.2.
Roland, dem die Pflegerin ängstlich den Arzt aus der Klinik ruft, ist von der Leiter gefallen.
29.3.
Roland, dem die Pflegerin äußerst beunruhigt den Arzt aus der Klinik ruft, ist von der Leiter gefallen.
29.4.
Roland, dem die Pflegerin äußerst ängstlich den Arzt aus der Klinik ruft, ist von der Leiter gefallen.

30.1.
Ferdinand, dem der Gärtner erschreckt den Hengst vom Bauernhof ausleiht, will nach Köln reiten.
30.2.
Ferdinand, dem der Gärtner fassungslos den Hengst vom Bauernhof ausleiht, will nach Köln reiten.
30.3.
Ferdinand, dem der Gärtner äußerst erschreckt den Hengst vom Bauernhof ausleiht, will nach Köln reiten.
30.4.
Ferdinand, dem der Gärtner äußerst fassungslos den Hengst vom Bauernhof ausleiht, will nach Köln reiten.

31.1.
Hans, dem der Bruder verletzt den Fisch aus der Zoohandlung aufdrängt, muss ihn abholen fahren.
31.2.
Hans, dem der Bruder angeschlagen den Fisch aus der Zoohandlung aufdrängt, muss ins abholen fahren.
31.3.

Hans, dem der Bruder äußerst verletzt den Fisch aus der Zoohandlung aufdrängt, muss ihn abholen fahren.
31.4.
Hans, dem der Bruder äußerst angeschlagen den Fisch aus der Zoohandlung aufdrängt, muss ins abholen fahren.

32.1.
Patrick, dem die Verlobte fasziniert den Hochzeitsfotografen aus Hamburg aussucht, holt beim Schneider die Brautjungfernkleider ab.
32.2.
Patrick, dem die Verlobte verliebt den Hochzeitsfotografen aus Hamburg aussucht, holt beim Schneider die Brautjungfernkleider ab.
32.3.
Patrick, dem die Verlobte äußerst fasziniert den Hochzeitsfotografen aus Hamburg aussucht, holt beim Schneider die Brautjungfernkleider ab.
32.4.
Patrick, dem die Verlobte äußerst verliebt den Hochzeitsfotografen aus Hamburg aussucht, holt beim Schneider die Brautjungfernkleider ab.

33.1.
Francesco, dem die Bäckerin begeistert den Fischverkäufer aus Neapel empfiehlt, fährt mit seiner Vespa zum Hafen.
33.2.
Francesco, dem die Bäckerin glücklich den Fischverkäufer aus Neapel empfiehlt, fährt mit seiner Vespa zum Hafen.
33.3.
Francesco, dem die Bäckerin äußerst begeistert den Fischverkäufer aus Neapel empfiehlt, fährt mit seiner Vespa zum Hafen.
33.4.
Francesco, dem die Bäckerin äußerst glücklich den Fischverkäufer aus Neapel empfiehlt, fährt mit seiner Vespa zum Hafen.

34.1.
Hubert, dem die Ärztin beleidigt den Psychiater in der Stadt nahelegt, hat ernsthafte Probleme.
34.2.
Hubert, dem die Ärztin gekränkt den Psychiater in der Stadt nahelegt, hat ernsthafte Probleme.
34.3.
Hubert, dem die Ärztin äußerst beleidigt den Psychiater in der Stadt nahelegt, hat ernsthafte Probleme.
34.4.
Hubert, dem die Ärztin äußerst gekränkt den Psychiater in der Stadt nahelegt, hat ernsthafte Probleme.

35.1.
Christian, dem der Onkel begeistert den Jagdhund aus Belgien ersteigert, lernt bald schießen.

35.2.
Christian, dem der Onkel glücklich den Jagdhund aus Belgien ersteigert, lernt bald schießen.
35.3.
Christian, dem der Onkel äußerst begeistert den Jagdhund aus Belgien ersteigert, lernt bald schießen.
35.4.
Christian, dem der Onkel äußerst glücklich den Jagdhund aus Belgien ersteigert, lernt bald schießen.

36.1.
Nico, dem die Kollegin fasziniert den Professor aus Cambrigde zeigt, studiert in England.
36.2.
Nico, dem die Kollegin verliebt den Professor aus Cambridge zeigt, studiert in England.
36.3.
Nico, dem die Kollegin äußerst fasziniert den Professor aus Cambrigde zeigt, studiert in England.
36.4.
Nico, dem die Kollegin äußerst verliebt den Professor aus Cambridge zeigt, studiert in England.

37.1.
Klaus, dem der Stiefvater erzürnt den Freund aus der Schule ausredet, ärgert sich über das Thema.
37.2.
Klaus, dem der Stiefvater aufgebracht den Freund aus der Schule ausredet, ärgert sich über das Thema.
37.3.
Klaus, dem der Stiefvater äußerst erzürnt den Freund aus der Schule ausredet, ärgert sich über das Thema.
37.4.
Klaus, dem der Stiefvater äußerst aufgebracht den Freund aus der Schule ausredet, ärgert sich über das Thema.

38.1.
Andi, dem die Partnerin verärgert den Brieffreund aus Frankreich verleidet, hat ihm sowieso schon lange nicht mehr geschrieben.
38.2.
Andi, dem die Partnerin ärgerlich den Brieffreund aus Frankreich verleidet, hat ihm sowieso schon lange nicht mehr geschrieben.
38.3.
Andi, dem die Partnerin äußerst verärgert den Brieffreund aus Frankreich verleidet, hat ihm sowieso schon lange nicht mehr geschrieben.
38.4.
Andi, dem die Partnerin äußerst ärgerlich den Brieffreund aus Frankreich verleidet, hat ihm sowieso schon lange nicht mehr geschrieben.

39.1.
Helmut, dem die Kommissarin erzürnt den Gefangenen aus der Haftanstalt übergibt, mag die Transporte nicht.
39.2.

Helmut, dem die Kommissarin zornig den Gefangenen aus der Haftanstalt übergibt, mag die Transporte nicht.
39.3.
Helmut, dem die Kommissarin äußerst erzürnt den Gefangenen aus der Haftanstalt übergibt, mag die Transporte nicht.
39.4.
Helmut, dem die Kommissarin äußerst zornig den Gefangenen aus der Haftanstalt übergibt, mag die Transporte nicht.

40.1.
Markus, dem die Dame entsetzt den Boten vom Amt zurückschickt, vergisst oft zu unterschreiben.
40.2.
Markus, dem die Dame erschrocken den Boten vom Amt zurückschickt, vergisst oft zu unterschreiben.
40.3.
Markus, dem die Dame äußerst entsetzt den Boten vom Amt zurückschickt, vergisst oft zu unterschreiben.
40.4.
Markus, dem die Dame äußerst erschrocken den Boten vom Amt zurückschickt, vergisst oft zu unterschreiben.

41.1.
Jochen, dem die Ehefrau amüsiert den Jongleur aus Düsseldorf bestellt, ahnt nichts von der Initiative.
41.2.
Jochen, dem die Ehefrau heiter den Jongleur aus Düsseldorf bestellt, ahnt nichts von der Initiative.
41.3.
Jochen, dem die Ehefrau äußerst amüsiert den Jongleur aus Düsseldorf bestellt, ahnt nichts von der Initiative.
41.4.
Jochen, dem die Ehefrau äußerst heiter den Jongleur aus Düsseldorf bestellt, ahnt nichts von der Initiative.

42.1.
Manfred, dem der Türsteher frustriert den Betrunkenen vor der Disko anvertraut, kennt sich mit solchen Fällen aus.
42.2.
Manfred, dem der Türsteher gefrustet den Betrunkenen vor der Disko anvertraut, kennt sich mit solchen Fällen aus.
42.3.
Manfred, dem der Türsteher äußerst frustriert den Betrunkenen vor der Disko anvertraut, kennt sich mit solchen Fällen aus.
42.4.
Manfred, dem der Türsteher äußerst gefrustet den Betrunkenen vor der Disko anvertraut, kennt sich mit solchen Fällen aus.

43.1.
Erich, dem die Sängerin überzeugt den Sportler aus Rom vorstellt, ist ziemlich schüchtern.

43.2.
Erich, dem die Sängerin entschlossen den Sportler aus Rom vorstellt, ist ziemlich schüchtern.
43.3.
Erich, dem die Sängerin äußerst überzeugt den Sportler aus Rom vorstellt, ist ziemlich schüchtern.
43.4.
Erich, dem die Sängerin äußerst entschlossen den Sportler aus Rom vorstellt, ist ziemlich schüchtern.

44.1.
Jonas, dem der Vater begeistert die Braut auf dem Stadtfest auswählt, mag keine Bevormundungen.
44.2.
Jonas, dem der Vater glücklich die Braut auf dem Stadtfest auswählt, mag keine Bevormundungen.
44.3.
Jonas, dem der Vater äußerst begeistert die Braut auf dem Stadtfest auswählt, mag keine Bevormundungen.
44.4.
Jonas, dem der Vater äußerst glücklich die Braut auf dem Stadtfest auswählt, mag keine Bevormundungen.

45.1.
Kevin, dem die Patentante verzaubert den Künstler aus Berlin bestellt, eröffnet bald seine Galerie.
45.2.
Kevin, dem die Patentante träumerisch den Künstler aus Berlin bestellt, eröffnet bald seine Galerie.
45.3.
Kevin, dem die Patentante äußerst verzaubert den Künstler aus Berlin bestellt, eröffnet bald seine Galerie.
45.4.
Kevin, dem die Patentante äußerst träumerisch den Künstler aus Berlin bestellt, eröffnet bald seine Galerie.

46.1.
Otto, dem der Galerist fasziniert die Malerin auf dem Empfang vorstellt, ist entzückt.
46.2.
Otto, dem der Galerist verliebt die Malerin auf dem Empfang vorstellt, ist entzückt.
46.3.
Otto, dem der Galerist äußerst fasziniert die Malerin auf dem Empfang vorstellt, ist entzückt.
46.4.
Otto, dem der Galerist äußerst verliebt die Malerin auf dem Empfang vorstellt, ist entzückt.

47.1.
Sebastian, dem der Schreiner amüsiert den Lehrling aus dem Möbelgeschäft anvertraut, ist eigentlich Zimmermann.
47.2.
Sebastian, dem der Schreiner heiter den Lehrling aus dem Möbelgeschäft anvertraut, ist eigentlich Zimmermann.
47.3.
Sebastian, dem der Schreiner äußerst amüsiert den Lehrling aus dem Möbelgeschäft anvertraut, ist eigentlich

Zimmermann.
47.4.
Sebastian, dem der Schreiner äußerst heiter den Lehrling aus dem Möbelgeschäft anvertraut, ist eigentlich Zimmermann.

48.1.
Egon, dem der Maurer schockiert den Hilfsarbeiter aus Weißrussland zurückschickt, kümmert sich schlecht um seine Angestellten.
48.2.
Egon, dem der Maurer geschockt den Hilfsarbeiter aus Weißrussland zurückschickt, kümmert sich schlecht um seine Angestellten.
48.3.
Egon, dem der Maurer äußerst schockiert den Hilfsarbeiter aus Weißrussland zurückschickt, kümmert sich schlecht um seine Angestellten.
48.4.
Egon, dem der Maurer äußerst geschockt den Hilfsarbeiter aus Weißrussland zurückschickt, kümmert sich schlecht um seine Angestellten.

D SRN training grammar

S : **NPnom VP eos |** *# comments*
{C_num, NPnom Nnom, VP V Vtrans} | *# number agreement main clause*
{C_num, NPnom Nnom, VP V Vditrans}; *# number agreement main clause*
NPnom : **Nnom RC |** *# noun phrase nominative*
{C_num, Nnom, RC SRC RPnom} | *# number agreement NPnom & RelPron SRC*
{C_num, Nnom, RC SRC V Vtrans} | *# number agreement NPnom & Verb SRC*
{C_num, Nnom, RC SRC V Vditrans} | *# number agreement NPnom & Verb SRC*
{C_num, Nnom, RC ORC ORCakk RPakk} | *# number agreement NPnom & RelPron ORCakk*
{C_num, Nnom, RC ORC ORCdat RPdat}; *# number agreement NPnom & RelPron ORCdat*
VP : **V indObj ADVP NPakk |**
{C_trans, V, indObj}; *# if transitive, add indirect object (NPdat)*
indObj : **NPdat | no**;
V : **Vtrans | Vditrans** ; *# verbs can be transitive or intransitive*
NPakk : **Nakk RC |** *# noun phrase accusative*
{C_num, Nakk, RC SRC RPnom} | *# number agreement NPakk & RelPron SRC*
{C_num, Nakk, RC SRC V Vtrans} | *# number agreement NPakk & Verb SRC*
{C_num, Nakk, RC SRC V Vditrans} | *# number agreement NPakk & Verb SRC*
{C_num, Nakk, RC ORC ORCakk RPakk} | *# number agreement NPakk & RelPron ORCakk*
{C_num, Nakk, RC ORC ORCdat RPdat}; *# number agreement NPakk & RelPron ORCdat*
NPdat : **Ndat RC |** *# noun phrase dative*
{C_num, Ndat, RC SRC RPnom} | *# number agreement NPdat & RelPron SRC*
{C_num, Ndat, RC SRC V Vtrans} | *# number agreement NPdat & Verb SRC*
{C_num, Ndat, RC SRC V Vditrans} | *# number agreement NPdat & Verb SRC*
{C_num, Ndat, RC ORC ORCakk RPakk} | *# number agreement NPdat & RelPron ORCakk*
{C_num, Ndat, RC ORC ORCdat RPdat}; *# number agreement NPdat & RelPron ORCdat*
RC : **SRC** (0.03) **| ORC** (0.06) **| no** ; *# 9% relative clauses,*
 # 1/3 SRC, 2/3 ORC (1/2 ORCakk, 1/2 ORCdat)
SRC : **RPnom NPakk ADVP indObj V |** *# subject extracted relative clause*
{C_trans, indObj, V}; *# if transitive, add indirect object (NPdat)*
ORC : **ORCakk | ORCdat**; *# object extracted relative clauses (accusative or dative)*
ORCakk : **RPakk NPnom ADVP NPdat Vditrans |** *# object extracted relative clauses accusative*
RPakk NPnom ADVP Vtrans |
{C_num, NPnom Nnom, Vditrans} | *# number agreement in RC*

```
          {C_num, NPnom Nnom, Vtrans};               # number agreement in RC
ORCdat :  RPdat NPnom ADVP NPakk Vditrans |          # object extracted relative clauses dative
          {C_num, NPnom Nnom, Vditrans};             # number agreement in RC
## terminal nodes
Nnom   :  der_pilot_NOM_SING | die_piloten_NOM_PLUR |
          der_kommandant_NOM_SING | die_kommandanten_NOM_PLUR;
Nakk   :  den_piloten_AKK_SING | die_piloten_AKK_PLUR |
          den_kommandanten_AKK_SING | die_kommandanten_AKK_PLUR;
Ndat   :  dem_piloten_DAT_SING | den_piloten_DAT_PLUR |
          dem_kommandanten_DAT_SING | den_kommandanten_DAT_PLUR;
Vtrans :  ueberrascht (0.05) | ueberraschen (0.05) | begeistert (0.2) | begeistern (0.2) |
          sieht (0.125) | sehen (0.125) | hoert (0.125) | hoeren (0.125);
Vditrans : zeigen | zeigt | empfehlen | empfiehlt;
ADVP   :  MOD ADV | "" ;                             # add an adverbial phrase in 50%
MOD    :  ausserst | "";                             # add a modifying adverb to an adverbial phrase in 50%
ADV    :  ueberrascht | begeistert |                 # ambiguous adverbs
          freudig | gluecklich;                      # unambiguos adverbs
# relative pronouns
RPnom  :  der_RP_NOM_SING | die_RP_NOM_PLUR;                  # nominative
RPakk  :  den_RP_AKK_SING | die_RP_AKK_PLUR;                  # accusative
RPdat  :  dem_RP_DAT_SING | denen_RP_DAT_PLUR;                # dative
no     :  "";                                                 # empty, no relative clause

###### constraints ######

C_num{                                               ### number agreement
                                                     ## singular, left hand side
        der_pilot_NOM_SING |                         # nouns nominative
        der_kommandant_NOM_SING |
        den_piloten_AKK_SING |                       # nouns accusative
        den_kommandanten_AKK_SING |
        dem_piloten_DAT_SING |                       # nouns dative
        dem_kommandanten_DAT_SING |
        der_RP_NOM_SING |                            # relative pronouns
        den_RP_AKK_SING |
        dem_RP_DAT_SING
        :                                            ## singular, right hand side
        ueberrascht | begeistert | sieht | hoert |   # transitive verbs
        zeigt | empfiehlt |                          # ditransitive verbs
        der_RP_NOM_SING |                            # relative pronouns
```

```
        den_RP_AKK_SING |
        dem_RP_DAT_SING;
                                                ## plural, left hand side
        die_piloten_NOM_PLUR |                  # nouns nominative
        die_kommandanten_NOM_PLUR |
        die_piloten_AKK_PLUR |                  #nouns accusative
        die_kommandanten_AKK_PLUR |
        den_piloten_DAT_PLUR |                  # nouns dative
        den_kommandanten_DAT_PLUR |

        die_RP_NOM_PLUR |                       # relative pronouns
        die_RP_AKK_PLUR |
        denen_RP_DAT_PLUR
        :                                       ## plural, right hand side
        ueberraschen | begeistern | sehen | hoeren |   # transitive verbs
        zeigen | empfehlen |                    # ditransitive verbs
        die_RP_NOM_PLUR |                       # relative pronouns
        die_RP_AKK_PLUR |
        denen_RP_DAT_PLUR;
        }

C_trans{                                        ### transitivity
        Vtrans : no;                            # if transitive verb, no indirect object
        Vditrans : NPdat;                       # if ditransitive verb, indirect object
        NPdat : Vditrans;                       # if indirect object, use ditransitive verb
        no : Vtrans;                            # if no indirect object, use transitive verb
        }
```