The Information Gathering Framework.
A Cognitive Model of Regressive Eye Movements during Reading

INAUGURAL-DISSERTATION

zur

Erlangung der Doktorwürde

des Fachbereichs

Germanistik und Kunstwissenschaften

der Philipps-Universität Marburg

vorgelegt von

Anna Fiona Weiß

aus Siegen

Marburg, im Oktober 2017

Vom Fachbereich 09: Germanistik und Kunstwissenschaften der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 04.10.2017.

Datum der Disputation: 14.11.2017

Dekan: Prof. Dr. Malte Hagener

Betreuerin / Erstgutachterin: Prof. Dr. Ina Bornkessel-Schlesewsky
University of South Australia
Adelaide, Australien

Zweitgutachter: Prof. Dr. Richard Wiese
Philipps-Universität Marburg
Marburg, Deutschland

SOLI DEO GLORIA

DANKSAGUNGEN

Mit großer Dankbarkeit blicke ich auf die letzten Jahre zurück. Ich bin überwältigt von der Hilfe, den zahlreichen Ermutigungen und praktischen Unterstützungen, die ich erfahren durfte. Besonders gefreut hat es mich, dass so viele Menschen ehrlich und aufrichtig Anteil an dieser Promotion mit ihren Höhen und Tiefen genommen haben. Dafür möchte ich mich hiermit einfach ganz herzlich bedanken. Solche Menschen zu kennen, ist ein großes Geschenk für mich.

An erster Stelle möchte ich mich bei Ina Bornkessel-Schlesewsky bedanken – meiner Doktormutter, wie man so schön sagt. Du hast mich für die Neurolinguistik begeistert, mich von Anfang an gefördert (obwohl ich als verstaubte Altphilologin für die rasanten Entwicklungen der modernen Neurowissenschaften wohl ziemlich ungeeignet erschien) und immer an mich geglaubt. Auch wenn die räumliche Distanz die Zusammenarbeit bisweilen erschwert hat, konnte ich mich immer auf Deinen äußerst kompetenten fachlichen Rat und Dein bedingungsloses Wohlwollen verlassen. Dadurch hast Du mich ermutigt, auch Dinge auszuprobieren, die ich mir selbst nicht zugetraut hätte. Danke, liebe Ina, für Dein Vertrauen und Deine Unterstützung!

I also thank Adrian Staub. I do not believe I ever told you how much I appreciated your advice and your support. You always had an open ear and an encouraging word for me. And in particular, you always responded incredibly fast to my emails! It has been a great pleasure for me to work with you. Thank you so much.

Franziska Kretzschmar danke ich für viele Jahre der treuen Unterstützung. Wie oft hast Du durch Dein unfassbares Wissen über Blickbewegungen meinen eigenen Blick bewegt und warst mir in langen Gesprächen über Fachliches und Nicht-Fachliches eine wertvolle Ansprechpartnerin. Danke, Franziska, für die viele Zeit und Mühe, die Du in mich investiert hast!

Aber auch bei vielen weiteren Personen möchte ich mich herzlich bedanken. Sarah Tune und Arne Nagels haben mich nicht nur in die (Un-)Tiefen der fMRT-Analyse eingeführt, sondern hatten auch immer ein offenes Ohr für mich. Auf Richard Wiese konnte ich mich in all den Jahren voll und ganz verlassen, und so haben wir nicht nur Chorauftritte, sondern auch viele organisatorische und universitäre Herausforderungen gemeinsam gemeistert. Ulrike Domahs als meine neue Chefin hat mir immer mit Rat und Tat den Rücken gestärkt und mir alle Freiräume geschaffen, die ich für das Fertigstellen meiner Arbeit brauchte.

Aber auch meinen übrigen (teilweise ehemaligen) Kolleginnen und Kollegen am IGS / DSA möchte ich herzlich danken, insbesondere (neben vielen weiteren) Kathrin Siebold, Stephanie Leser-Cronau, Manuela Lanwermeyer, Karin Wiese und Alexander Dröge, die immer wieder ein Mittagessen sowie Freuden und Sorgen mit mir geteilt haben. Wie schön ist es, in einem so angenehmen Institut mit so vielen netten Menschen arbeiten zu dürfen!

Auch all den anderen Menschen, die auf die eine oder andere Weise an meinen Forschungsprojekten beteiligt waren, möchte ich hiermit herzlich danken. Das waren Jona Sassenhagen, Phillip Alday, Katerina Kandylaki, Joshua Levy, Nikhil Garg, Wesley Albright, Atreyi Mukherji, Chhing Tiv, Hannah Simes, R. Muralikrishnan, Davide Laneri, Christina Lubinus, Jens Sommer, Mechthild Wallnig, Peer Herholz und Miriam Burk.

Abschließend möchte ich mich ganz herzlich bei meiner Familie und meinen (Glaubens-) Geschwistern bedanken. Eure Gebete und Eure Gemeinschaft haben mich über all die Jahre getragen. Ihr standet immer hinter mir, habt mich ermutigt und wart mir wertvolle Ratgeber. Also, I would like to thank my sisters and brothers from abroad, in particular Annelise and Duncan Wood, not only for proofreading my dissertation but also for the wonderful time I spent with you in Adelaide! Als meine große Familie habt Ihr immer wieder meine Prioritäten zurechtgerückt und mich daran erinnert, dass wir einen genialen Vater im Himmel haben, für den einfach „keine Sache zu wunderbar ist“ – und wenn ich etwas in den letzten Jahren lernen durfte, dann das!

Marburg, im Oktober 2017

Anna Fiona Weiß

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

INTRODUCTION

There is no other cultural invention that has had such an incredible impact on our society as reading. The idea that simple combinations of lines and marks may represent a complex system of language, enabling people to express demanding trains of thought, as well as deep feelings or controversial opinions, independent of space and time, is profound. Indeed, there is no longer any need for the recipient of the linguistic message to be in the same room at the same time, and there is not even a need for the recipient to live in the same century. This concept which at first glance appears trivial has allowed revolutionary new ways of human communication.

Although the language system itself remains the same in both written and spoken modalities, it seems comprehensible that the opportunity to encode the linguistic message in a written form has also influenced the way the language system is used. Written language, for example, is generally more conceptualized as a monolog, whereas the dialog is the dominant form of spoken language. Spoken language also includes references to the immediate situation in which the linguistic utterance is expressed (like *here*, *there*, *this morning* and so on), while written language has to be more explicit and precise due to the absence of a shared context. Crucially, written language (with the exception of chat communication) seems to be more complex and homogenous with regard to sentence structure, colloquial usage, dialect-isms and so forth than spoken language (see Dürscheid, 2012, for an overview of conceptual differences between written and spoken language).

From a cognitive-oriented language processing perspective, however, the specific characteristics of written compared to spoken language can be summarized by the following three points:

a) The linguistic input for the language processing system is encoded in visual units. This means that the eyes (or the hands for Braille lettering, respectively) and the operating principles of the oculomotor system are prerequisites for a successful processing of written language.
b) The visual input units are distinctive, at least in most of the world's writing systems. In contrast to the continuous phonological input of spoken language, the segmentation of the visual input of written language into distinctive visual units and words that have a certain spatial extension seems much easier, especially for people who begin to learn a foreign language.

c) The linguistic input is (usually) permanent, at least over a certain amount of time. Thus, many aspects of the reading process are under the (conscious and unconscious) control of the reader himself, who in turn decides for how long he[1] will read a certain part of the sentence. Crucially and in clear contrast to spoken language, the reader is able to reread parts of the linguistic input.

Whereas all of these aspects have been subject to extensive research in the last decades, the dominant research questions have focused upon basic letter and word identification and integration processes. Besides many open questions, it is now widely understood how visual units are identified, linked to single words, associated with lexical meanings and integrated into the syntactic structure (see, e.g., reading models like E-Z Reader: Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006; or SWIFT: Engbert, Nuthmann, Richter, & Kliegl, 2005; that are able to simulate human reading behavior on the basis of computational models). Also, the determinants of these processes such as frequency and predictability or attentional control have been examined in a variety of experimental settings (for an overview see Rayner, 1998; 2009).

However, although the opportunity to reread parts of the linguistic input is one of the basic characteristics of reading, this phenomenon has never been in the broader focus of reading research. For example, whereas it is well known that difficulties in sentence comprehension may cause rereading (e.g., Frazier & Rayner, 1982), it is still unclear, what factors exactly trigger rereading and to what extent the rereading process is under linguistic control. Furthermore, it is not even clear what role rereading plays in the development of an effective reading strategy that is fast and leads to a sufficient sentence interpretation at the same time.

The present work shall examine rereading processes during sentence interpretation in the light of cognitive and neural processing mechanisms. Thus, the two central research questions are:

a) What is the role of rereading processes in the context of successful sentence interpretation?
b) To what extent are the underlying cognitive and neural processes under control of the reader's language processing system?

[1] For simplicity reasons we will refer to the reader only with the masculine pronoun but this always includes female readers as well and should not be assumed as a form of valuation.

In order to address these questions, we will first review the basic findings reported in the literature with regard to rereading processes and their implementation into current models of eye movement control. Based on these considerations, we will outline a new account, called the *Information Gathering Framework*, which aims to overcome some weaknesses of the former models and to provide a general tool for our understanding of rereading processes during sentence interpretation.

In the second part of this thesis, we will present empirical evidence from two experiments that were performed in order to test hypotheses that are derived from the *Information Gathering Framework*. The first experiment focuses on the influence of specific tasks evoked by different question types on reading processes, especially on rereading. The second experiment examines the neural correlates of rereading using a concurrent eye tracking and functional magnetic resonance imaging technique.

PART I: THE COGNITIVE MECHANISMS OF READING AND REREADING

Most people read even difficult passages of text without any trouble. Only in certain cases are skilled readers aware that in order to extract meaning from a sequence of visual units, reading involves a variety of single processing steps that work in a highly automatic manner: Reading involves processes of letter and word recognition, recall of meanings from the lexicon as well as integration of words into phrases, sentences and texts, based on syntactic and other principles. Finally, an interpretation of the phrases, sentences and texts has to be computed that also incorporates information about the context, the genre and the author itself, before an evaluation of what has been read can take place. In addition, all of these processes can be further divided into numerous sub-processes.

This chapter provides a short summary about the basic principles of reading. However, the focus is set on the phenomenon of rereading which is discussed in more detail. Based on prior research, it is first reviewed what we know about regressive eye movements and their function in the process of reading. In particular, we ask the questions a) what triggers rereading? and b) what part of the sentence is read again? We will also see how current models of eye movement control during reading answer these questions. In a second step, we will use these findings to develop a new reading model, the so-called *Information Gathering Framework*, which in particular aims to capture the interplay between increased reading times and the decision to reread.

1. Reading – An alternate pattern of fixations and saccadic eye movements

During reading it seems that the eyes smoothly slide through the sentence. But this is just an illusion. In fact, the eyes show an alternate pattern of almost stable rest phases (so called *fixations*) and very rapid jumps (so called *saccades*). Due to the very fast motion of the eyes during a saccade, new information can only be acquired during a fixation (Matin, 1974).

Saccades are necessary in order to place the fovea, the part of the eye's retina with the highest acuity, on the part of the stimulus that the reader wants to see clearly. Outside the fovea region (that covers about 2 degrees in the center of vision), the acuity in the parafoveal region (1–5 degrees away from the center of vision) decreases significantly and in the peripheral region (everything beyond the parafoveal region) the ability to distinguish details, colors or shapes is very limited (Rayner, 2009).

Before the eyes move to a new location, the movement has to be planned and executed. Empirical findings show that the time needed to choose the target location in the visual field and to initiate a saccade (so called *saccade latency*) is of about 175–200 ms (Becker & Jürgens, 1979; Rayner, Slowiaczek, Clifton, & Bertera, 1983). The time needed to actually move the eyes, depends on the distance moved and takes about 20–35 ms in reading (Rayner, Schotter, Masson, Potter, & Treiman, 2016; Abrams, Meyer, & Kornblum, 1989; Rayner, 1978).

During silent reading, the average fixation duration for experienced adult readers is about 225–250 ms and saccades in alphabetic writing systems have a mean length of 7–9 letters and cover about 2 degrees of the visual field (Rayner, 2009). Although this is still a subject of debate, most researchers agree that decisions about how long to fixate a word and when to move the eyes to the next word, are not subject to an automatic mechanism that scans the sentence in always the same manner, but are to a great extent under the cognitive control of the reader (Rayner, Liversedge, White, & Vergilino-Perez, 2003; Reingold, Reichle, Glaholt, & Sheridan, 2012). Nevertheless, the considerable variability in fixation durations reflects such things as characteristics of the text (font, light / dark contrast), linguistic difficulty (predictability, word frequency, ambiguity), properties of the reader (reading skill, age, linguistic knowledge) or task (proofreading, skimming, reading for comprehension; see Rayner, Schotter, Masson, Potter, & Treiman, 2016; Rayner, 2009; Rayner, 1998).

There is no direct linkage between the single fixation duration and the processing of a certain word because not every word is fixated during reading. In English, function words are skipped about 65% of the time whereas content words are only skipped about 15% of the time. That is the case because there is a clear correlation between the word's length and its probability to be skipped: Short words tend to be skipped more often than longer words do (Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996). In addition, when the word length increases it is also more likely that a word is fixated by the reader more than once (McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; McDonald & Shillcock, 2004). Words that are very frequent and / or highly predictable from the prior context also tend to be skipped. However, that a word has not been fixated does not mean that it has not been processed at all. All current models of reading assume a processing of skipped words, at least to some extent (Rayner et al., 2016).

Crucially for the current purpose is the finding that the eyes do not always move forward through the text, but that they move backwards about 10–15% of the time in skilled readers (Rayner, 2009; see Figure 1 for an example of a typical eye movement pattern when reading a sentence).

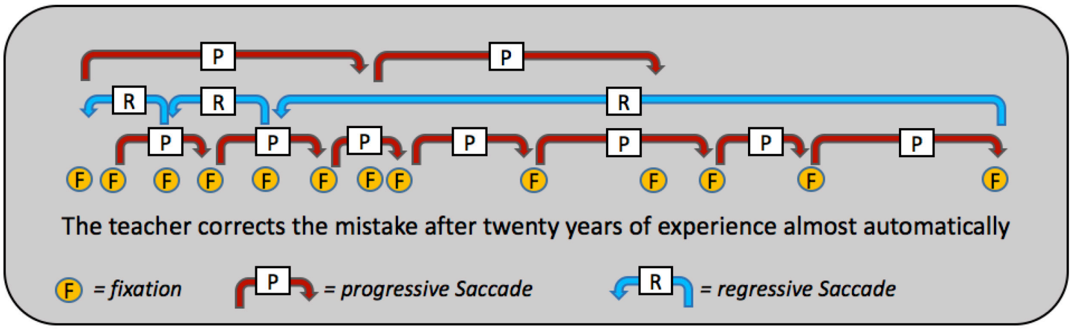

Figure 1: Typical eye movement pattern during reading a sentence (a word-by-word translation from German). The eyes first walk through the sentence, then regress from the end and start rereading.

Whereas some of these backward saccades just re-fixate the current word (so called *intra-word regressions*) which happens quite often for long words with 7 or more letters (Rayner et al., 2016), most of the backward saccades move back to a previous word (so called *inter-word regressions* or just *regressions*). There is evidence that inter-word and intra-word regressions have to be viewed as functionally distinct because they seem to be differently influenced by factors such as word frequency or word length (Vitu & McConkie, 2000). In addition, intra-word regressions do not result in rereading because the target position has not been fixated earlier (i.e., they take place because of a suboptimal initial fixation). However, the general opportunity to reread earlier parts of the text (and to move the eyes against the reading direction by a regression) is crucial for written language processing (see Introduction). For spoken language processing, by contrast, the listener (usually) is not able to decide how long he listens to a certain part of the auditory input or whether he wants to listen again to earlier parts of the message.

In the following we will focus on these regressive eye movements and review the most important findings from previous research with regard to the two central questions:

a) What triggers a regressive eye movement?
b) What determines the landing position of a regressive eye movement?

2. Regressive eye movements during reading

In his comprehensive review on eye movements in reading, scene perception, and visual search, Keith Rayner points out that *"Regressions are not particularly well understood because it is difficult to control them experimentally"* (Rayner, 2009). Of course, in light of the fact that regressive eye movements are a fundamental part of the human reading mechanism, it seems quite surprising that more than 30 years of reading research have not been able to sufficiently answer the basic questions, for instance, about the causes and targets of regressions. One reason might be that it is indeed difficult to manipulate regressions in experimental settings, although the developing of new statistical methods allows for the use of e.g. unbalanced designs. It seems, however, that the main reason for our poor understanding of regressions lays in another fact, namely, that regressions find themselves caught between two different research traditions.

In the past decades reading has been studied extensively with a wide range of experimental manipulations. But there are (at least) two different approaches in reading research: The first approach focuses on reading itself, the mechanisms of fixations and saccade generation as well as their interaction with attention and perceptual span. The second approach uses reading as a tool to understand language processing, particularly focusing on higher levels of language like syntax and semantic processing. Most of the findings of the language approach are not specific for reading, based on the assumption that language processing works – besides the differences in the input itself, i.e., visual vs. auditory input – identical in both modalities (for a discussion see Bader, 2015).

Regressive eye movements, however, are somewhere between these two approaches. On the one hand, they are specific for reading and occur only in the visual modality. Thus, they are an elementary component of our understanding of the basic principles and mechanisms of reading. On the other hand, regressions are not just generated by low level features of the visual input such as word length, or linguistic properties like word frequency or predictability. Rather, they occur mainly in the context of difficulties with the semantic or syntactic interpretation of the sentence (Rayner et al., 2016). This means consequently that a reading-oriented understanding of regressive eye movements, without taking into account the principles of higher order language processing, is just as doomed to failure as a language-orientated understanding without taking into account the principles of reading.

It is therefore important for a deeper insight into the role of regressive eye movements for sentence interpretation to consider both aspects: the mechanisms of reading and the principles of language processing.

2.1 What triggers a regressive eye movement?

In the context of psycholinguistic research, eye movements are viewed to correlate with cognitive processes of language interpretation. This idea that is known as the *eye mind hypothesis* was first developed by Just and Carpenter (1984) and laid the basis for studying eye movements in order to learn something about the mechanisms of language processing. In this regard, especially two reading measures are assumed to reflect difficulties in language processing: increased (single and combined) fixation duration and a higher probability of regressive eye movements. Although it seems apparent from many previous studies that regressions reflect higher order problems in sentence interpretation as in garden path sentences (see, for example, Frazier & Rayner, 1982), it is unclear, however, what exactly causes a regressive eye movement and what just leads to increased fixation durations. Let us thus first consider some basic differences between these two reading measures before we turn to a more detailed discussion of regressions.

During a fixation, the eyes are not able to acquire new information outside the visual field. From a perspective that focuses on problem solution, this means that the parser can only deal with information currently available to the processing system (which includes, of course, information that can be retrieved from short-term memory). Thus, increased fixation durations should reflect a response to problems that may be solved by the currently available information (like difficulties of lexical access). Regressions, by contrast (as progressions, of course), enable the parser to acquire new information and, in the case of regressions, information that has already been processed, at least to some extent. Thus, regressions should reflect a response to problems that may not be solved by the currently available information (like misinterpretations of the syntactic structure in the case of ambiguities). This means that regressions provide information in order to solve comprehension difficulties.

On a first glance, this distinction between increased fixation duration and probability of regressions seems trivial but in the following we will take this idea as a starting point for our understanding of regressive eye movements. To begin with, we propose the working hypothesis that gathering information is the driving principle for triggering a regression in sentence reading. Let us therefore first consider some empirical findings and explanations discussed in the literature that might help us to make this concept more specific.

The previously described working hypothesis assumes that regressive eye movements are under the control of higher-order language processing mechanisms and not just executed by bottom-up information of the visual input. This is, in fact, quite a controversial claim and there is indeed evidence that this is not always the case. For example, eye movements do not always reach their intended target position because the oculomotor system under- or overshoots the saccade. This happens frequently e.g. after return sweeps to the next line of text which tend to fall short of the beginning of the line (Andriessen & De Voogd, 1973). In response to this motor error, a correcting eye movement is performed, and, in the case of an overshoot, this results in a saccade to the left. These correcting saccades occur frequently and are assumed to work without any linguistic control (see Bicknell & Levy, 2011). In addition, most of them are very short and fall within the same word (intra-word regressions) whereas only a small number actually land on a previous word. Since these regressions do not result in rereading of previous sentence material (the preceding fixation tends also to be very short and the regression usually lands on a position somewhere between the last two fixation positions, i.e. not on a position that has been fixated earlier), they can hardly be compared with inter-word regressions that occur during sentence reading.

Besides the discussion of corrective saccades, the idea has been brought forward that regressive eye movements may reflect difficulties or failures in word identification (Bouma, 1978; Bouma & De Voogd, 1974; Pollatsek & Rayner, 1990). In this case, it is hypothesized that the eyes are directed backwards to the word that has not been fully identified. One of the most influential studies that investigated the general pattern of regressions with regard to word identification difficulties was carried out by Vitu and McConkie (2000). Vitu and McConkie analyzed an eye-tracking corpus consisting of 4 adults reading a classical novel and identified quintuples of three fixations and two saccades illustrated in the following scheme:

Fixation 1	Saccade 1	Fixation 2	Saccade 2	Fixation 3
origin word	always progressive, may skip words	destination word	progressive or regressive	target word

Thereafter, they examined the probability and length of regressions (saccade 2) depending on a broad variety of factors. Two results of the study are of particular interest for our present purpose and were interpreted by the authors as evidence for the *word identification hypothesis*: First, inter-word regressions occurred more frequently in cases where the word prior to fixation 2 had been skipped and saccade 1 was more than 6 characters long. Second, if the origin word was long and low in frequency, the probability for a regression to the origin (=target) word was also increased.

Although Vitu and McConkie do not concretize the *word identification hypothesis*, the basic idea is in line with our working hypothesis proposed in the beginning: Regressions are triggered in order to compensate for deficits of information, in this case of information about the identity of a particular word. This also means that regressive eye movements are assumed to reflect linguistic processing and not just a spatial re-orientation or other low-level evaluations. In addition, the results for skipped words fit well with the proposed working hypothesis: Skipped words are assumed to be processed parafoveally but not to the extent fixated words are (Rayner et al., 2016). Since primarily short words tend to be skipped, it seems also plausible that especially in the case of longer words the parafoveal intake of relevant information for word identification is more difficult. Thus, the available information about a skipped word's identity is reduced and in order to compensate for this deficit the reader regresses to the skipped word.

However, the findings of the influence of the origin word's properties on regression probability seems more challenging for our working hypothesis: Why did the reader not just fixate longer when he was not able to identify the word in first pass reading? We are not able to answer this question satisfactorily yet, but some explanations are possible here. The most plausible reason could be that the eyes move forward to word *n+1* before the processing of word *n* is completed, an idea that is incorporated in several reading models (see, e.g., E-Z Reader: Reichle et al., 2006; Reichle, Warren, & McConnell, 2009; or SWIFT: Engbert et al., 2005). In this case, the reader recognizes on word *n+1* that the information about word *n* is not sufficient enough to identify the word and in response to this lack of information, a regressive eye movement is performed. But note that the pattern observed by Vitu and McConkie only accounts for a more or less small number of regressions because they only took quintuples where the second fixation landed on the origin word. In addition, the effect of frequency and word length was not very strong. Thus, it is reasonable to conclude that the characteristics of the origin word only play a minor role, and that problems of lexical access should primarily result in increased fixation durations. On the other hand, these considerations make it difficult, of course, to generate clear predictions about the expected eye movement pattern in response to problems of lexical access. But one promising approach might be to analyze the regression probability depending on first pass times, to see if shorter first pass times would lead to an increased regression probability. This analysis, however, is not possible with the data provided by Vitu and McConkie.

A third category of explanations assumes that regressions reflect difficulties in higher-order language processing like syntactic and semantic integration (Bouma & De Voogd, 1974; Just & Carpenter, 1980). There is large evidence from sentence reading studies showing that semantically or syntactically anomalous sentences lead to increased regression probability, especially at the point or right after the difficulties become apparent (see Rayner & Pollatsek, 1989, for a review). With this regard, a particular focus was set on garden path sentences (see, e.g., Frazier & Rayner, 1982) where an initial interpretation of an ambiguous phrase turns out to be wrong later in the parsing process. In order to get the correct interpretation, the reader has to reanalyze the sentence structure by detecting the source of the error and by checking for alternative interpretations. There is an intensive debate how reanalysis exactly takes place (see Fodor & Ferreira, 1989, for a review), but a prerequisite for successful reanalysis seems to be that the relevant information about the identity, relative order or inflectional markings of words that occur earlier in the sentence is available to the parser. It could be the case that this information can be retrieved from short-term memory because it normally has been processed earlier (at least to some extent) and that reanalysis just takes place during a fixation (covert reanalysis), but in the case this is not possible (or not sufficient), the missing information has to be acquired by the execution of a regressive eye movement (overt reanalysis).[2,3] Thus, difficulties in syntactic and semantic integration provide a reasonable account why regressions are triggered in sentence reading and also fit well with the working hypothesis proposed in the beginning: Regressive eye movements are executed in order to gather additional information.

Yet however convincing this approach might appear, comprehension difficulties cannot account for all regressions. Particularly, there is evidence that regressions do also occur in sentences that do not contain a garden path (for a discussion, see von der

[2] Note that missing information not necessarily means that the information is not available at all. It could also be that the confidence about the information is low (see Bicknell & Levy, 2010).

[3] Note that garden path sentences can also be processed auditorily so that regressions (and additional information) cannot be a general prerequisite for reanalysis.

Malsburg & Vasishth, 2013). Of course, it is possible that comprehension difficulties may also arise in syntactically and semantically well-formed sentences, at least in a random manner (e.g., due to individual lack of attention). But the concept of comprehension difficulties as the driving principle for regressions is further challenged by the findings of a general increase of fixation duration and regressions at the end of a sentence, known as "sentence wrap-up effects" (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000; Hirotani, Frazier, & Rayner, 2006). These regressions are largely independent of the presence of an anomalous or garden path structure and therefore cannot be attributed to the reanalysis processes or comprehension difficulties in general. Rather, it seems that the reader is checking his sentence interpretation by rereading former linguistic material. However, to what extent this rereading mechanism is under linguistic control and triggered by certain sentence structures remains unclear.

With regard to our general understanding of regressions, we therefore have to slightly modify our working hypothesis proposed in the beginning. Instead of limiting regressions to problem solution, we propose that regressive eye movements occur whenever the processing system needs to refresh information to the left of fixation in order to fulfill the task at hand. This task can either be to develop a coherent interpretation of the sentence or to recall the sentence structure or anything completely different. This means, conversely, that the task at hand should also have an impact on the decision to execute a regressive eye movement.

Summarizing so far, it seems that regressions are not just triggered by bottom-up information provided by the visual input features. Rather, regressive saccades during sentence reading seem to be driven by a linguistic control (with the exception of corrective regressions). But it is unclear, however, to what extent this linguistic control influences the shape of the regression pattern and which linguistic processes exactly cause regressions. Whereas none of the approaches discussed above can account for all types of regressions, we propose the idea that the unifying function underlying all regressions during sentence reading might be to gather relevant additional information that is needed in the course of sentence interpretation. This may, but need not, include information that is required in order to solve comprehension difficulties. In this regard, regressions should reflect primarily higher-order language processing like syntactic or semantic integration of linguistic material into the sentence structure, but could also in some cases be a response to difficulties in word identification. In any case, regressive eye movements seem to play a crucial role in developing a coherent and reliable sentence interpretation in cases where first pass reading is not sufficient. Thus, the opportunity to reread former parts of the linguistic input provides a useful compensation strategy which enables the parser to deal with information deficits: This is an opportunity that is not present in auditory sentence processing and might explain why readers can understand complex sentences better than listeners do (Miller & Smith, 1989).

2.2 Regressions in the context of current models of eye movement control

In the last decades, the development of computational simulation models that can account for eye movement control during reading has had a great impact on our understanding of basic reading processes. But how do they explain the role of regressions in the reading process? We will now discuss the basic assumptions of the two most influential models, E-Z Reader 10 (Reichle et al., 2009) and SWIFT (Engbert et al., 2005), with regard to their modelling of regressive eye movements. Afterwards we will review (to our knowledge) the only model of eye movement control in reading that explicitly focuses on regressive eye movements. This model was proposed by Bicknell and Levy (2010).

2.2.1 The E-Z Reader 10 Model

The E-Z Reader model (Reichle et al., 1998; Reichle, Pollatsek, & Rayner, 2006) was first developed to account for the interplay between lexical processing, attention allocation and saccadic programming during reading and had essentially nothing to say about higher level language processing. However, the latest version of the model, E-Z Reader 10 (Reichle et al., 2009), now also tries to explain the interaction between 'post-lexical processing' and eye movement control.

One of the basic assumptions of E-Z Reader is that saccade programming is decoupled from attention shifting so that the processing of a word is not strictly correlated to its fixation duration. Thus, after an early stage of lexical processing, the *familiarity check*, the oculomotor system is programming a saccade to the next word *n+1* while the second stage of lexical processing, the *completion of lexical access*, is completed on the current word *n*, before the attention also shifts to word *n+1*. In E-Z Reader 10, the former model was expanded by a post-lexical integration stage where the word *n* is integrated into higher-level representations like the syntactic structure or the discourse model. It is assumed that if the integration of word *n* fails before word *n+1* is identified, this results in comprehension difficulty. In addition, the failure of integration causes both an attention shift and a regressive eye movement *"back to the point at which the difficulty became evident (i.e., word n), as opposed to some earlier sentence location"* (Reichle et al., 2009, p. 6).

Note that the authors state explicitly, that *"the integration stage […] is a placeholder for a deeper theory of postlexical language processing during reading. Our goal in* *including this stage is therefore quite modest: to provide a tentative account of how […] post-lexical variables might affect readers' eye movements."* (p. 6). In other words, the E-Z Reader 10 model assumes that problems with postlexical integration cause regressive eye movements and the model does fairly a good job in simulating reading behavior. But whereas difficulties with postlexical integration per se undoubtedly play a crucial role for triggering a regression (see chapter 2.1), the model provides no description about the precise nature of these problems due to the absence of a detailed theory of language processing.

Importantly and in contrast to our working hypothesis, the primary goal of a regression is assumed to be a redirection of attention, namely a redirection of attention back to the point of integration failure. According to the model, this point of integration failure is restricted to word *n*. There are two problems arising from this view. First, the redirection of attention to word *n* allows for re-processing (and maybe deeper processing) of word *n*, i.e. the available information for solving the problem is in fact information about the identity of word *n*. In the case word *n* cannot be integrated in the sentence structure, the possible solutions are those that can be retrieved from the word's identity (e.g., lexical ambiguity). This may be helpful in some cases but in the majority of occasions it is much more relevant to figure out what actually causes the problem. And the cause of the problem often does not match the region where integration fails and the problem becomes apparent (c.f. the ambiguous and disambiguation region in garden path sentences). In this regard, it seems a more efficient strategy in the case of postlexical integration difficulties to redirect attention to the cause of the problem instead to the point where integration actually fails (which might, of course, coincide in some cases).

Second, the model can only account for regressions targeting word *n* and, in addition, only for postlexical integration difficulties, which is a simplification in both ways. On the one hand, regression target locations show a more complex distribution pattern (see chapter 2.4) but on the other hand, postlexical integration difficulties cannot account for all types of regressions, as has been discussed earlier.

In sum, although E-Z Reader 10 is a very helpful tool in our understanding of basic reading processes, its basic assumptions about the role of regressions in the context of reading are underspecified in some regards. The reason for that lies particularly in the lack of an elaborated theory of language processing that allows for more precise predictions about how integration problems may be solved.

2.2.2 The SWIFT Model

The SWIFT model, proposed by Engbert and colleagues (Engbert, Longtin, & Kliegl, 2002; Engbert et al., 2005), is a further highly advanced model that aims to mathematically simulate the control of eye movements during reading on both psychological and neurophysiological properties. Engbert and colleagues computed model fits on the basis of individual words and explicitly tried to do so without an advanced model of language processing, but on the basis of simplified rules for word recognition and mechanisms for saccade programming. The authors propose that a general mechanism, based on a few core principles, underlies all types of saccades.

In contrast to the serial word processing in the E-Z Reader model, the SWIFT model assumes that multiple words are processed in parallel while saccades are generated autonomously, per default targeting the next word. This automatic saccade programming can be canceled if a new saccade is initiated during a first labile level of saccade programming (which happens, e.g., in skipping). Importantly, saccade programming and targeting is primarily driven by word recognition which influences the activation of potential saccade targets on a saliency map (the activation field). A maximal activation on the activation field is also related to the word's processing difficulty, which in turn includes the parameters frequency and predictability. Thus, words that are low in frequency and predictability have a high activation on the activation field and therefore a high probability to become the target of the next saccade.

According to the minimalization approach of the SWIFT model, word recognition is the driving principle for all types of saccades which also includes regressive eye movements. This means that regressive eye movements are assumed to be triggered by an incomplete word recognition. In this case, the eyes are re-directed to the word where the recognition failed. Because words that are skipped in first pass reading tend not to be processed (and recognized) in full detail, the model also assumes an increased probability for a skipped word to become the target of a regression. Although the authors show that incomplete word recognition may account for regression probabilities in experimental data, they conclude that *"more constraints are needed to estimate regression probabilities"* (Engbert et al., 2005, p. 792). Note that regression probability in the context of the SWIFT model does not mean the probability to regress on a certain word but the probability that a particular word is the target of a regression.

Taken together, the SWIFT model expands the basic assumptions of the *word identification hypothesis* discussed in chapter 2.1. by incorporating them into a general model of eye movement control. Thus, the basic ideas are in line with our working hypothesis assuming that regressions aim to gather additional information. Although some principles like the prediction of an increased re-inspecting probability for skipped words align well with the empirical findings (see Vitu & McConkie, 2000), there are some limitations and problems with the SWIFT model in the modelling of regressive eye movements which we will discuss below.

The first question addresses the concept of incomplete word recognition. To recognize the identities of the words in a sentence is a prerequisite for sentence interpretation and comprehension. The identity of a word provides information (among many others) about its meaning and semantic neighborhood as well as its word class and the possible functions in the sentence structure and allows furthermore for predictions about other entities in the sentence and their specific role in the sentence structure. Since word recognition covers a broad variety of information, it is thus reasonable to assume that word recognition is not always complete and that information is missing. But at the same time this raises the question if word recognition could ever be completed at all. Consider the wide range of alternatives in the sub characteristics of a word that interact with each other: There is often more than one word class possible, more than one meaning, more than one semantic neighborhood, more than one prediction. To maintain all of these alternatives equally until a final decision can be made by additional evidence later in the sentence (or text), would probably lead to excessive memory demands. On the other hand, if incomplete word recognition triggers a regressive eye movement and if we assume that most (if not all) words are incompletely recognized, this should result in an enormous number of regressive eye movements. It is thus very important to specify this concept of incomplete word recognition more precisely:

a) What does complete word recognition actually include?
b) Does incomplete word recognition necessarily lead to a regression or does this just happen in the case the incomplete word recognition leads to difficulties in word integration?
c) Why has the word recognition been incomplete? What factors matter?

A second question addresses the target position of regressions: The SWIFT model claims that "*a regression can occur because of unfinished lexical access before the corresponding region of text is left*" (p. 790). Thus, a forward saccade is triggered although the word recognition has not been finished. In response, a regression to this particular word is performed in order to finish lexical access. However, according to the SWIFT model only regressions to the immediately preceding word may occur. But how, then, does the model account for regressions that target words earlier in the sentence? Note again that the SWIFT model does focus on the probability that a word becomes the target of a regression and assumes that this happens independently of the characteristics of the word from which a regressive eye movement is launched.

Finally, although the SWIFT model incorporates the concept of predictability as a factor that determinates the target position of saccades, the model can account for higher-order language processing only in a limited manner. Of course, information about the identity of a word is one of the core tasks of all linguistic processes and problem solutions. But diagnosing the source of syntactic processing difficulties, for instance, and providing an efficient repair strategy involves more than just rereading the word which was incompletely recognized. It may include re-structuring of former linguistic material on the basis of word order regularities and syntactic principles that cannot be just derived from single word properties. Thus, to detect the point in the sentence that contains the relevant information for the linguistic process under consideration requires a precise hypothesis about the possible error sources based on lexical and non-lexical information. In addition, as has been discussed earlier, word recognition cannot account for all types of regressions.

These considerations show that assessing the role of regressive eye movements during reading is a very challenging task and has to take into account the mechanisms of language processing as well as the empirical findings of eye movement control and attention shift. Thus, although the SWIFT model has many strengths, it is underspecified in many ways (as is the E-Z Reader model) and does not provide a comprehensive and fully convincing explanation of regressive eye movements during reading.

2.2.3 The Model of falling confidence[4]

Due to the limitations of the E-Z Reader and the SWIFT model with regard to regressive eye movements, Bicknell and Levy proposed a new model of eye movement control that aims to overcome the weaknesses of the former models (Bicknell & Levy, 2010). At the core, it is assumed that a word identification process never is completed. Thus, *"it is pos-*

[4] The term „Model of falling confidence“ was never used by Bicknell and Levy, they just called their approach "A rational model of Eye Movement Control in Reading" (Bicknell & Levy, 2010). But due to the great number of models of eye movement control in reading, which might cause confusions, we named it the "Model of falling confidence", for short "FC model" (see also Bicknell & Levy, 2011).

sible that later parts of a sentence can cause a reader's confidence in the identity of the previous regions to fall" (Bicknell & Levy, 2010, p. 1170) which triggers a regressive eye movement in order to get more visual information about the previous region.

According to the framework, the model generates distributions over possible identities of the sentence, based on its language model. During a fixation, the noisy visual input is used to update the model's beliefs by a Bayesian likelihood term and by the language model. Thereupon, the model selects an action which could either be to continue fixating, to trigger a saccade or to stop reading the sentence before the cycle repeats.

A simple control policy is assumed to decide between actions, which works on the basis of two thresholds: The first value defines the threshold for a character to remain fixated. The second value defines the threshold for an (already processed) character on a leftward position to be fixated again (by a regression). Thus, the model allows to independently modulate the control policy with regards to processing depths (i.e., increased fixation durations) and regression probability which determines the speed and accuracy of the model. It is hypothesized that *"for any given level of speed and accuracy achieved by a non-regressive policy, there is a faster and more accurate policy that makes a faster left-to-right pass but occasionally does make regressions."* (Bicknell & Levy, 2010, p. 1174).

In order to test this hypothesis, Bicknell and Levy simulated the reading behavior of the model on a corpus containing 33 English sentences. The results suggest that making regressions indeed provides a useful strategy when confidence about previous words falls, and leads to a higher speed and better accuracy. Thus, falling confidence seems to provide a rational explanation for triggering a regressive eye movement.

As is apparent from the model description, the FC model fits well with the working hypothesis proposed in the beginning and offers a clear mathematical description of how such an account can be integrated into a simulation model. Furthermore, it takes the basic ideas of the SWIFT model but revises its assumptions with regards to incomplete word recognition by one essential modification: The problematic concept of "*incomplete word recognition*" that implies the existence of "*complete word recognition*" is replaced by the idea that word identification is never completed and that additional evidence later in the sentence may cause that additional information about a word's identity is required. Although this concept is still very vague, it resolves some weaknesses of the SWIFT model and provides a unifying account for regressive eye movements that is in accordance with the basic findings of regressions and respects characteristics of reading as well. But whereas the underlying principle of the FC model seems very convincing, the actual modelling of this principle in a computational model is again limited in several ways and raises further questions.

The major weakness of the model is that it has never been tested on human data. Bicknell and Levy took the model to simulate regression behavior on English sentences, but they just compared the efficiency of different reading strategies by adjusting the thresholds for the control policy and measured the resulting reading speed and accuracy in different simulations. Thus, it is completely unclear if the model is able to account for human reading behavior and if the assumptions of the model have any real-world reliability. The main focus of the modelling was set on the question whether a reading strategy that focuses on a high reading speed with occasionally making regressions is more efficient in terms of reading speed and accuracy than a more careful and slower reading strategy without making regression. This question was answered by theoretically motivated simulations and the authors concluded that the first strategy is more efficient.

Another limitation of the model is its underlying language model that is based on bigram frequency only. This includes that solely word *n+1* may cause confidence about the identity of word *n* to fall which in turn means that regressions are assumed to always be initiated on word *n+1* and to target word *n*. This view, however, is problematic for the following reasons: First, regressions do not always target the word prior to the current fixation, as is apparent from many empirical studies (see chapter 2.4 for a discussion). Second, the bigram focus is not able to explain a wide range of regressions that occur in response to higher-order language processing difficulties. In particular, it could be possible that word *n+1* causes the confidence about the identity of a word much earlier in the sentence to fall (and not just of word *n*). Consider, for example, garden path sentences where the ambiguous region is separated by several intervening words from the disambiguation region. Here it would be less useful to reread the disambiguation region or the region right before the disambiguation region when processing difficulties appear. Rather, in order to solve the problem (by reanalyzing the sentence), the source of the error has to be identified and an alternative interpretation of the ambiguous phrase has to be adopted. But the FC model cannot account for such instances nor does it provide a reasonable explanation (see for a similar argumentation the discussion of the E-Z Reader model in chapter 2.2.1).

Besides these considerations, note that the concept of falling confidence is (theoretically) not restricted to difficulties in language processing because it is unspecified with regard to the evidence that causes low confidence (which could also be a specific task, for example). Whereas this expands the theoretical validity of the framework, the ability of the actual model formalization for covering higher-order language processing phenomena is restricted. Again, the reduced language model can only partially account for difficulties that arise due to word order variations, syntactic ambiguities or semantic anomalies where often more information than just the bigram frequency of two consecutive words is needed to detect and to resolve the information deficit (see, e.g., garden path sentences). In addition, the model is rather vague with regard to the linguistic processes that cause low confidence and the question how a regression (and increased confidence) might help to solve the problem.

In sum, all three models of eye movement control discussed here (E-Z Reader 10, SWIFT, FC model) provide some important ideas how regressions can be integrated in the reading process. And whereas they differ with regard to the formalization of the mechanisms, they all agree that regressions are triggered by linguistic factors that include more than just basic word properties like single word frequency or word length. Rather, regressions are assumed to be executed in the context of several words, where some further evidence may reveal that an initial word identification process is not sufficient enough for building a coherent sentence interpretation. But none of the three models provides a unifying account that can cover the whole range of regressive eye movement patterns during sentence reading. An important factor might be that all of the three models are very restricted with regard to the linguistic processes they assume to trigger regressions.

2.3 Regressions and sentence comprehension

Before we will turn to the second important factor for our understanding of regressive eye movements, namely, the selection of regression targets, we will shortly deal with another question related to the role of regressions in sentence comprehension: If regressive saccades play a crucial role for developing a coherent sentence structure in reading (which is implicit in all of the accounts discussed above), do regressions then indeed lead to a better comprehension?

Schotter and colleagues (Schotter, Tran, & Rayner, 2014) addressed this question in a clever experiment. They had participants read ambiguous garden-path sentences and their unambiguous counterparts while all words to the left of the current fixation were replaced with an x-mask. Thus, the participants were able to regress to prior parts of the sentence but this regression did not provide any useful information. The results showed that comprehension significantly dropped down in the trailing-mask condition compared to normal reading. Remarkably, this was true for both ambiguous and non-ambiguous sentences, which again indicates that regressions do not exclusively provide information for a reanalysis but also for other difficulties in sentence interpretation.

In addition to the overall finding that the opportunity to regress supported comprehension, Schotter and colleagues also found evidence that actually making a regression had no significant effect on comprehension, i.e. regressions did not lead to significantly better comprehension results compared to cases where the reader did not regress. They interpreted their results as suggesting that *"when readers make a regression, they do so to improve failed comprehension and that regression improves their understanding to be equivalent to its level in cases in which they do not need to reread"* (Schotter et al., 2014, p. 6).

More recently, Metzner and colleagues (Metzner, von der Malsburg, Vasishth, & Rösler, 2016) compared sentence comprehension of free-reading and word-by-word presentation in a concurrent ERP / eye-tracking study. The results show that, besides the finding that regressions in the syntactically and semantically violated sentences seem to be associated with a P600 effect, accuracy in the judgement task improved when reading naturally compared to the word-by-word presentation. A more fine-grained analysis revealed, however, that the benefit was only visible when the eyes actually made a regressive saccade. When the readers did not regress, comprehension accuracy for the violated conditions was as low as or even lower than in the word-by-word presentation.

Thus, Metzner and colleagues concluded that their results suggest *"that the freedom to control viewing times alone does not improve comprehension"* and that *"the added benefit of natural reading really is the ability to revisit earlier material"* (Metzner et al., 2016, p. 20).

In sum, both experiments provide evidence that regressions play a crucial part in the comprehension of violated or ambiguous sentences. Nonetheless, they differ with regard to the benefit of a single regression: Whereas Schotter et al.'s results point to the conclusion that the opportunity for regressions improves comprehension overall, but actually making a regression supports comprehension only in a limited manner, Metzner et al. provide evidence that not only the opportunity for, but making a regression itself benefits comprehension.

It is not fully clear where these differences come from. One possibility for the lack of finding an overall benefit for regressions in Metzner et al.'s experiment may be the comparison with word-by-word presentation. In contrast to the hardly noticeable trailing mask in Schotter et al.'s experiment, this method could have led the participants to adopt a totally different and compensating reading strategy because they knew that the standard reading routine (flexible reading times, regressions) is not possible.

On the other hand, the use of very different sentence structures and manipulations could have had an influence on the benefit of making a regression. Whereas it is possible that some readers were just not able to parse the English garden path sentences in Schotter et al.'s experiment correctly (independently of making a regression or not), the violated German sentences in Metzner et al.'s experiment were less difficult to parse. In addition, the comprehension questions (well-formed vs. not well-formed) did not require a deep analysis of the sentence. Thus, it could be possible that in the non-regression trials in Metzner et al.'s experiment some readers just did not notice the violation and therefore did not regress. If the readers regressed, however, this helped them to detect the problem. In addition, the task itself (checking for well-formedness) may have encouraged the reader to read more carefully and to examine the relevant regions by an additional regression.

Very recently, another experiment was carried out by Christianson and colleagues (Christianson, Luke, Hussey, & Wochna, 2017, experiment 1). They also examined the relationship of rereading and comprehension accuracy when reading English garden path sentences with a reduced relative clause (GP RR, like 1) compared to local coherence structures also containing a reduced relative clause (LC RR, like 2).

(1) *The player tossed the ball interfered with the other team.*

(2) *The other team interfered with the player tossed the ball.*

Christianson and colleagues hypothesized that regressions may have two different functions. They may either lead to a reanalysis of and a "recovery" from an initially wrong parsing decision or they may function to confirm a previous parsing decision. Accordingly, the authors predicted that if regressions are used to confirm the interpretation of previous sentence material, the rereading of LC RRs should lead to more accurate results than the rereading of GP RRs since in the latter rereading will only confirm a misparse of the sentence which results in misinterpretation. By contrast, if regressions are used for recovery, then a regression in GP RRs should lead to better accuracy.

The results of the study, however, showed no correlation of regressions into pre-critical regions and accuracy. Thus, as in the experiment conducted by Schotter and colleagues, actually making a regression did not improve accuracy. In addition, the analysis revealed very low accuracy results. For the ambiguous GP RRs the participants answered less than 25% of the sentences correctly, and for the LC RRs the accuracy was about chance. Together with shorter reaction times for GP RRs, Christianson and colleagues interpreted their findings as a hint for confirmatory rereading. Thus, the readers did not reanalyze the sentence, even when rereading the sentence, and came finally up with the wrong interpretation.

These findings are of course striking in several ways. They in particular raise the question what function regressions have in sentence interpretation if making a regression does not lead to better comprehension. But note again that both types of ambiguous sentences used in the experiment were very hard to understand. In addition, the comprehension questions (e.g., *Did the player toss the ball*?) also had to be first comprehended correctly (which is a challenging task given the absence of any plausibility information and the reliance on syntactic computations only) and required an exact syntactic processing of the thematic roles in the stimulus sentences. Thus, even for the unambiguous control structures the mean accuracy was clearly below 75%[5] which could indicate general difficulties in answering the comprehension questions. Together with the findings from Schotter et al. and Metzner et al. this suggests that regressions are a helpful mechanism in sentence comprehension but that their benefit is limited by the linguistic

[5] Note that the exact values are not reported in the paper.

knowledge of the reader. In other words, even a regressive eye movement would be useless if the reader does not have the ability to deal with the linguistic problem.

2.4 The target position of regressive eye movements

In the previous chapters, we have seen that although regressive eye movements occur frequently in reading, none of the approaches proposed in the literature can account for all types of regressions. Even the current models of eye movement control during reading are not able to accurately predict or simulate the occurrence of regressive eye movements in sentence reading, primarily because regressions are not restricted to lexical processing of single words. Nonetheless, the basic assumptions of the FC model, that are in line with our working hypothesis proposed in the beginning, provide a good starting point for our understanding of regressive eye movements in reading.

Furthermore, although the actual formalization of the FC model is too restricted in order to account for the full pattern of regressions during reading, the basic ideas of the FC model and our working hypothesis contain a very important implication: If regressions aim to gather information, then the information itself should play a crucial role in our understanding of regressive saccades. In other words, the cause of a regression should be related to its intended target position.

In the following we will review the basic findings with regard to target positions of regressive eye movements. Although the question whether such a target position exists or not and how we could define this target position is a crucial aspect of regressions, the empirical evidence is limited to only a few studies, primarily focusing on garden path sentences. Thus, the question of target positions is still far from being answered.

2.4.1 The selective reanalysis hypothesis

The first experiment examining the target position of regressive saccades during reading was conducted by Frazier and Rayner and dates back to the 1980s (Frazier & Rayner, 1982). Frazier and Rayner assumed that eye movements are directly driven by linguistic processes and that the decision when and where to regress is under strict linguistic control. Accordingly, eye movements should provide an insight into the parsers reanalysis processes during sentence comprehension.

Based on the finding that garden path sentences lead to an increased probability of regressions out of the disambiguating region, Frazier and Rayner proposed three different opportunities how the parser could deal with these processing difficulties: a) The parser reads the entire sentence again (*forward reanalysis hypothesis*), b) The parser systematically proceeds backward and tries out alternatives (*backward reanalysis hypothesis*), c) The parser regresses to a position where he expects the source of the error (*selective reanalysis hypothesis*).

In their experiment, Frazier and Rayner let participants read English garden path sentences like (3) while the participants' eyes were monitored.

(3) Since Jay always jogs a mile and a half seems like a very short distance to him.

An inspection of the regression proportions revealed that 51% of all regressions in garden path sentences were initiated from the disambiguating region (*seems*) where the initially preferred sentence interpretation turned out to be wrong. In addition, 53% of all regressions initiated in the disambiguating region and beyond ended in the ambiguous region (*a mile and a half*) which indicates that the eyes were directly moved to the point where the erroneous analysis was developed. These findings support the *selective reanalysis hypothesis* that predicts that the parser regresses to the region where he expects the source of the error.

Nonetheless, Frazier and Rayner also found evidence that although regressions provide a typical strategy in order to deal with comprehension difficulties, readers do not always regress when faced with garden path sentences. Rather, in some cases comprehension difficulties just increased fixation durations. This finding has not been discussed by the authors in more detail but points to the opportunity that reanalysis might be covert, i.e., without triggering a regressive eye movement. In addition, this suggests that the availability of information that is needed to solve processing problems might depend on individual memory capacities or different sentence material (see also Lewis, 1998).

However, two further results of the study have to be mentioned with regard to the *selective reanalysis hypothesis*. First, Frazier und Rayner pointed out that beyond the dominant pattern of selective reanalysis, there was an increased probability of regressions initiated from the very end of the sentence and targeting the sentence beginning. Thus, the readers reread the entire sentence after finishing it for the first time. These sometimes called "sentence wrap-up" effects occur frequently in reading, as discussed earlier, but they question the general validity of the *selective reanalysis hypothesis*. Thus, Frazier and Rayner conclude that *"it is not obvious how regressions to the beginning of the sentence should be interpreted"* (Frazier & Rayner, 1982, p. 204).

The second point also addresses the general validity of the *selective reanalysis hypothesis*. The results show that although selective reanalysis seems to be the dominant pattern for the target selection of regressions, the regressions nonetheless show a relatively high variance with regard to their landing sites. Thus, if the *selective reanalysis hypothesis* is indeed the leading principle of regressive eye movements, then this account also has to explain why so many regressions obviously fail to reach their intended target position.

The *selective reanalysis hypothesis* is further questioned by the fact that the ambiguous region in the sentences was directly followed by the disambiguation region. Thus, the possibility cannot be ruled out that the eyes in certain cases just by default regressed to the word directly to the left of the currently fixated word because it is the *"smallest possible regression"* (Mitchell, Shen, Green, & Hodgson, 2008), independently of whether this word contained the source of the error or not.

The major limitation of Frazier and Rayner's experiment, however, was the lack of statistical evidence due to the overall small number of regressions. Thus, twenty years after Frazier and Rayner's famous experiment, Meseguer, Carreiras and Clifton further examined the *selective reanalysis hypothesis*, using locally ambiguous sentences in Spanish (Meseguer, Carreiras, & Clifton, 2002).

In the experiment, forty-four native speakers of Spanish were asked to read sentences like (4) while their eyes were tracked.

(4) *El professor dijo que los alumnos*
se levantaran del asiento cuando los directores entraron / entraran en la clase.

[The teacher said that the students
had to stand up from their seats when the directors came / come into the room]

In these sentences, the verb *entraron / entraran* disambiguates the ambiguous adverb *cuando* to be either high attached to the matrix verb *dijo* (in the case of *entraron*) or to be low attached to the second verbal phrase *se levantaran* (in the case of *entraran*). The results revealed a preference for the low VP2 attachment, indicated by increased second pass reading times and a higher probability of regressions for the VP1 attached sentences in the post-disambiguation region. In addition, statistical analysis provided significant differences between the two sentence types with regard to the target position of regressions: Regressive eye movements that were initiated in the post-disambiguation region targeted more frequently the matrix verb *dijo* and the adverb *cuando* in the sentences with VP1 attachment than in sentences with VP2 attachment. Hence, the authors interpreted their findings as additional evidence for the *selective reanalysis hypothesis*.

But although the results are generally in line with the findings of Frazier and Rayner, they raise some further questions and problems. The first problem is a methodological issue and refers to the sentence structure. As mentioned earlier, there is a general tendency to regress from the end of a sentence, independently of comprehension difficulties. The results of Frazier and Rayner show that these end-of-sentence regressions have to be functionally dissociated from regressions initiated in order to revise a misleading analysis of the linguistic material because the former tend to target the beginning of the sentence instead of a specific error source. In Meseguer at al.'s experiment, however, the sentence's final region coincided with the post-disambiguating region, so that end-of-sentence regressions can hardly be distinguished from regressions due to reanalysis (see also Mitchell et al., 2008, for a discussion).

The second issue refers to the first pass reading times in the disambiguation and post-disambiguation region. Whereas Frazier and Rayner reported increased fixation durations on the sentence region where the wrong analysis became apparent, Meseguer and colleagues found no differences between the two sentence types with regard to fixation durations in the disambiguation and post-disambiguation region. They interpreted this lack of an effect as a hint that the participants did not reanalyze the sentence covertly (i.e., without triggering a regressive eye movement). But together with the possible end-of-sentence regressions, this raises the question if all participants were equally garden pathed and if in fact all participants reanalyzed the sentence. In addition, the comprehension questions did not probe the attachment side of the sentences, so that it remains open with which sentence interpretation the participants finally came up with.

The third question addresses the reanalysis itself. Although Meseguer and colleagues provided a detailed description of how a sentence like (4) has to be reanalyzed in order to get the intended interpretation (i.e., attaching the adverb *cuando* to the matrix verb *dijo* instead of the second verb phrase *se levantaran*), they did not propose a clear hypothesis how this reanalysis is carried out. In other words, they did not specify the exact information (i.e., the portion of the sentence) that is needed in order to reanalyze the sentence. Remarkably, the results showed not only one but two preferred target regions: The matrix verb *dijo* and the adverb *cuando*. Meseguer and colleagues discussed this finding in their paper but were not able to give a satisfying answer. From the perspective of the *selective reanalysis hypothesis*, however, this raises the question to what extent a linguistically motivated single target region for regressions can be defined, even if the underlying language processing mechanisms are well studied.

In sum, although both experiments provide some support for the assumption that regressions are under linguistic control and are directed to the portion of the sentence where the parser expects the source for the error, they both suffer from methodological as well as interpretive problems. Particularly, the *selective reanalysis hypothesis* is restricted to regressions caused by comprehension difficulties in garden path sentences and therefore cannot account for all kinds of regressive eye movements. Thus, some doubts remain that regressions and their target positions are indeed driven by linguistic factors.

2.4.2 The time out hypothesis

In contrast to the strict *selective reanalysis hypothesis* claiming for ambiguous sentences that "*regressive eye movements will return directly to the ambiguous phrase"* (Frazier & Rayner, 1982, p. 188), Mitchell, Shen, Green and Hodgson (2008), however, proposed the idea that regressive eye movements just may reflect some kind of cognitive–inhibition mechanism. This so called *time out hypothesis* assumes that "*the function of the system is nothing more than that of postponing new input*" (Mitchell et al., 2008, p. 269). Thus, although the linguistic parser triggers a regression, the authors predict no linguistic guidance of the landing positions of regressive eye movements, they rather assume a more random distribution, largely characterized by short-range local saccades.[6]

In order to test these two conflicting approaches, Mitchell and colleagues conducted two experiments. The first experiment examined English garden path sentences like (5).

(5) *While those men hunted(,) the moose that was sturdy and nimble hurried into the woods and took cover.*

[6] Note that the *time out hypothesis* (as well as the *selective reanalysis hypothesis*) assumes that (at least a proportion of) regressions are triggered by the linguistic system and may thus reflect higher-order processing difficulties. But in contrast to the *selective reanalysis hypothesis* the *time out hypothesis* predicts no destination-linking of the landing positions driven by the linguistic system.

In these sentences, inserting a comma after the sub clause verb *hunted* would prevent the reader from being garden pathed. In addition to the punctuated and unpunctuated conditions, Mitchell and colleagues varied the position of the disambiguating region *hurried* from being the last word in the upper row (as shown above) or the first word of the lower row. This manipulation allowed them to distinguish linguistic effects from layout effects.

The results revealed strong evidence for layout-effects with, e.g., an increased number of regressions on the first line's final word and only a small number of regressions targeting the line above, independently of the linguistic manipulation. On the other hand, linguistic properties had an impact on regression behavior as well. But due to the small number of regressions launched from the disambiguating region (87 in total) and the corresponding lack of power as well as the potential effects of the physical appearance of the comma, it was difficult to draw reliable conclusions from the results.[7]

In a second experiment, Mitchell and colleagues took the punctuated and unpunctuated sentences with a late line break of experiment 1 (like the example given in 5) and constructed further sentences of the same type but with a late misanalysis area (as shown in example 6).

(6) *One sole hiker spotted that while those men hunted(,) the moose hurried into the woods and took cover.*

Based on the reanalysis accounts they predicted that according to the *selective reanalysis hypothesis* regressive eye movements launched from the disambiguating region *hurried* should target the words 4 (*hunted*) or 6 (*moose*) in the early misanalysis condition (example 5) and respectively the words 9 or 11 in the late analysis condition (example 6).

The results again were hard to interpret primarily due to the lack of power. However, the landing site distribution of regressions from the disambiguation region (*hurried*) did only provide a hint for an increased fixation of the ambiguous regions. But in the examination of all fixations falling within the regression sequence launched from the disambiguation region (and not just the first fixation), Mitchell and colleagues were able to show that indeed the target position of regressions primarily focused on the area of

[7] Note also that it is not really valid to leave out the comma in these types of sentences (see e.g., Staub, 2007).

words 4 and 6 in the early misanalysis condition and on word 9 and 11 in the late misanalysis condition, respectively.

Mitchell and colleagues interpreted their findings as clear evidence against the *time out hypothesis*, at least as a general principle[8], because the syntactic manipulation had a clear impact on the distribution of landing sites of regressive eye movements. Instead, the results point to the conclusion that the target positions of regressions are in fact tightly coupled with linguistic control according to the *selective reanalysis hypothesis*. But at the same time, as Mitchell and colleagues stated, the results do question the very strict assumption of the *selective reanalysis hypothesis* that the eyes are *directly* moved to the source of the error. Rather, "*regression guidance is characterized not by immediate and precise deployment but by tentative, step-by-step approaches to the target area*" (Mitchell et al., 2008, p. 285).

The results of Mitchell and colleagues thus confirm findings that could have been already observed (but were not explicitly discussed) in the experiments of Frazier & Rayner as well as Meseguer and colleagues, namely, that the landing positions are linguistically controlled but that there is still a high variance in the distribution of landing sites. But note that all the studies reviewed so far (Frazier & Rayner, 1982; Meseguer et al., 2002; Mitchell et al., 2008) were focusing on regressions triggered by the disambiguation region in garden path sentences. The results found here have to be considered with caution because they may be hardly transferred to other types of regressive eye movements.

2.4.3 Scanpath signatures in syntactic analyses

More recently, the degree of coupling of regressive eye movements and linguistic guidance was further investigated by von der Malsburg and Vasishth in two articles (von der Malsburg & Vasishth, 2011; 2013).

Based on the results of the three experiments discussed above (Frazier & Rayner, 1982; Meseguer et al., 2002; Mitchell et al., 2008), von der Malsburg and Vasishth proposed that the focus on single fixations might not be very informative with regard to the *pattern* of regressions that result when reanalysis begins. Thus, they developed a method that allowed them to analyze the scanpath of regressive eye movements during sentence

[8] Mitchell and colleagues leave it open whether at least some regressions are guided by a mechanism proposed by the *time out hypothesis*.

reading and to quantify the similarity of these patterns. In order to investigate the reanalysis processes in garden path sentences, they used this new approach to re-analyze the data from Meseguer et al.'s study.

In their scanpath analysis, von der Malsburg and Vasishth were able to identify three different reading patterns that occurred after making a regression out of the disambiguating (*entraran / entraron*) or post-disambiguating region. The first pattern (A) contained a long regressive saccade to the beginning of the sentence and subsequent rereading of the sentence. The second pattern (B) was characterized by a regression to the beginning of the sentence without rereading[9], and the third pattern (C) consisted of a short regression that targeted the disambiguating material. The analysis revealed that only pattern A (rereading) occurred significantly more often in the garden path conditions whereas patterns B and C were equally distributed over the two sentence types. Thus, von der Malsburg and Vasishth concluded that rereading is the preferred strategy when reanalyzing the sentence. In addition, the results suggest individual differences between readers so that for each reading pattern some readers were found who had a preference for this strategy.

The detected rereading pattern (A) stands in clear contrast to the predictions of the *selective reanalysis hypothesis* and questions the findings of Meseguer et al. who provided evidence for increased re-inspections of the matrix verb *dijo* (region 2) and the adverb *cuando* (region 6). Von der Malsburg and Visishth interpreted this contradiction in the way that the reported regressions of Meseguer et al. targeting regions 2 or 6 were just diverged in the course of the following fixations and did not build own patterns. In addition, the lack of findings of increased regression rates into the sentence initial region by Meseguer et al. is explained by short-falling regressions on their way to the sentence beginning.

Thus, von der Malsburg and Vasishth summarize that their results provide little evidence for the *selective reanalysis hypothesis* and point instead to a strategy comparable to the *forward reanalysis hypothesis* also outlined by Frazier and Rayner (1982) and where the reader starts reading the whole sentence again from the beginning in or-

[9] This pattern probably was a result of a missing fixation cross at the beginning of the line so that the readers anticipated the presentation of the new sentence. In the follow-up study, a fixation cross was used and subsequently the reading pattern disappeared (see von der Malsburg & Vasishth, 2011; 2013).

der to reanalyze it. On the other hand, the results do also contradict the *time out hypothesis* that assumes a random walk through the sentence because there was evidence for a more targeted regression process. Hence, von der Malsburg and Vasishth conclude that the results rather suggest that *"saccade programming is indeed influenced by low-level properties of the text while a loosely coupled linguistic system is guiding the overall shape of the trajectory."* (von der Malsburg & Vasishth, 2011, p. 124).

However, because the post-disambiguation region in the Meseguer et al. study was the sentence-final region as well, the results could have been interfered by regressions that were triggered by processes beyond the initial parsing of the sentence. Thus, von der Malsburg and Vasishth (2013) conducted a follow-up study using Meseguer et al.'s material but made three substantial adjustments: a) They extended the sentence final region to distinguish sentence wrap-up effects from parsing effects, b) they included a third, non-ambiguous condition in order to examine the interaction of regressions and parsing strategies, and c) they controlled for differences in working memory by performing an operation span test.

As in the former experiment, von der Malsburg and Vasishth identified three reading strategies that appeared after a regression out of the disambiguation or post-disambiguation region was made, from which two were identical to the ones observed earlier: pattern A (rereading) and pattern C (checking the disambiguating material). In addition, a third strategy was found that contained a short regression from the disambiguation region back to the pre-disambiguation region. Again, only the rereading pattern occurred more often in the high-attachment condition and this effect was more pronounced for readers with a high working memory capacity. Von der Malsburg and Vasishth interpreted their findings as additional evidence that rereading is the dominant strategy when reanalyzing a sentence. However, we think that the results lead to three more important conclusions which we will discuss below.

First, the authors argue that predictions about landing sites of regressions are directly linked to the underlying parsing model. Frazier and Rayner, for example, based their hypotheses on the assumptions of the garden path model and expected comprehension difficulties on the disambiguation region because the parser incorporates the phrase "*a mile*" as an object into the main-clause verb "*jogs*" according to the late closure principle. Whereas there is indeed some evidence that readers parse this sentence in the way the garden path model proposes, the parsing strategies of the Spanish sentences used by Meseguer et al. (and von der Malsburg and Vasishth), for example, are less clear.

Von der Malsburg and Vasishth discussed the predictions of five parsing models (garden-path model, construal model, good-enough parsing, parallel model, unrestricted race model) with regard to the sentences they used in their experiment and showed that all of them should result in different eye movement patterns. In particular, the models differ with regard to expected parsing difficulties which are assumed to trigger a regressive eye movement. Since regressions did occur in both garden-path and non garden-path sentences, this raises the question if rereading is restricted to reanalysis or if reanalysis is required in either sentence type. This, of course, should influence the expected regression pattern. Thus, in order to understand regressions in the context of sentence parsing it seems necessary to control more carefully with which interpretation the parser comes up with. We think this is a very important issue.

Second, although von der Malsburg and & Vasishth were also focusing on comprehension difficulties that showed up in first pass reading times and tried to identify the underlying parsing principles (they interpreted their findings as evidence for the good enough approach), they treated regressions out of the (post-)disambiguation region as reanalysis processes. Especially, these reanalysis processes were assumed to be carried out in order to find the correct interpretation of the sentence. But the very low accuracy in the comprehension questions for high-attached sentences (58%) shows that a large number of participants failed to parse these sentences correctly although some of them regressed to reread earlier parts. Since von der Malsburg and Vasishth did not examine the relationship of regression patterns and accuracy, three important questions remain open:

a) Were the readers able at all to diagnose the error in their interpretation correctly, independently of their reanalysis behavior? This ability is a prerequisite for the *selective reanalysis hypothesis* but due to the very poor accuracy results it seems likely that while many readers came up with comprehension difficulties in first pass reading, they were not able to localize the problem.
b) Was there a preference to use a certain regression strategy when readers knew the source of the error compared with when they just had no idea? Again, the very poor accuracy results could indicate that rereading was the preferred strategy because many readers were completely overburdened and were not able to develop hypotheses on what might have went wrong.

c) Was there a reading strategy that was more efficient and led to better comprehension (and accuracy results)? Maybe rereading did occur most frequently but rarely helped to find the correct interpretation whereas another pattern was more efficient but due to the small number of high-proficient readers it did not show up often.

Finally, the results of the first analysis (von der Malsburg & Vasishth, 2011) showed that readers may have a preference for one reading strategy although the reason for this preference was not specified. In addition, the results of experiment 2 revealed that the reader's working memory capacity might influence the parsing decisions (i.e., that low-capacity readers tended to leave the attachment-decision open till the disambiguating information was apparent[10]) and led to different reading strategies, too. These findings suggest that individual differences between readers could influence parsing decisions and reading strategies. Thus, we should take into account that it might be hard to find one guiding principle for all readers that allows to predict the scanpath signature of regressive eye movements.

[10] This also fits in nicely with findings from the ERP literature (see, e.g., Gunter, Wagner, & Friederici, 2003).

3. A new approach: The *Information Gathering Framework*

After we have reviewed the basic findings on regressive eye movements in sentence reading, we can sum up by saying that the picture they reveal is not very consistent: Although there is little doubt that regressions are linguistically driven, the strength of this linguistic control remains unclear. Whereas approaches like the *selective reanalysis hypothesis* assume a very strong linguistic guidance on triggering and landing sites of regressions, other experimental findings have shown that this hypothesis cannot fully account for the variability in target positions of regressions. Also, it seems implausible that regressive eye movements are just a response to difficulties in integrating linguistic material since regressions do not only occur in the context of processing problems. Thus, none of the approaches discussed so far may capture the full pattern of regressive eye movements during reading and an elaborated model of regressive saccades is still a major desideratum, even after many decades of reading research.

The goal of the current thesis is therefore to formulate a new framework that may provide a general tool for our understanding of regressive eye movements, without limiting it to a small range of linguistic phenomena. As a starting point, we will use the *model of falling confidence* proposed by Bicknell and Levy (2010). But instead of focusing on theoretical considerations about reading strategies, the current aim is to develop a realistic model of human reading behavior, which means that the model should be able to cover findings from the existing literature as well as make further testable predictions about reading behavior. This, however, requires some substantial modifications in the architecture of the FC model, so that we will call the new account the *Information Gathering Framework*.

Note that in contrast to the *model of falling confidence*, the *Information Gathering Framework* is not incorporated into a computational model as yet that allows for simulating reading behavior. Instead, the *Information Gathering Framework* takes into account more cognitive and linguistic properties of eye movement control than the former model does. But the current considerations should be used by future research to combine these two approaches and to develop a computational version of the *Information Gathering Framework* as well.

3.1 The architecture of the *Information Gathering Framework*

In the following, we will first briefly outline the basic ideas of the model, before laying out the new assumptions of the *Information Gathering Framework* and clarifying the modifications from the FC model in more detail. Afterwards, we will apply this framework to several empirical findings reported in the literature before we formulate a couple of predictions based on the model.

3.1.1 Model overview

The main properties of the *Information Gathering Framework* can be summarized by the following six assumptions:

a) The confidence in each word's identity is described by the confidence level. The confidence level is computed by matching predictions about incoming material with the lexical representation of a word.
b) The lexical representation of a word is viewed as an infinite bundle of features which takes time to be retrieved and which varies among individuals.
c) The focus of attention (i.e., the area within the confidence levels are computed in parallel) is restricted to two words.
d) There are two different thresholds for the confidence level that cause an action: The forward threshold defines the confidence level that is needed to trigger a progressive eye movement, whereas the backward threshold prevents a regression.
e) There are two different scenarios that cause a regressive eye movement: First, if the confidence level falls under the forward threshold after the eyes have already moved to the next word, and second, if the backward threshold is not reached before the confidence level of the next word reaches the forward threshold.
f) There are also two different scenarios as to how a regression target is selected: Either by re-computing the confidence levels within the perceptual span or by using experience-based strategies.

We will now explain these assumptions in more detail below, in particular with regard to the FC model.

3.1.2 Information gathering as the unifying principle of regressive eye movements

It is possible that regressions simply reflect some increased processing demands due to problems in the course of sentence interpretation (e.g., as proposed by Mitchell et al., 2008) and that the eye movement itself contributes little or nothing to the solving of that problem. In this case, we would expect a random or very restricted distribution of regression landing sites, preferentially targeting the immediately preceding word. Since this claim is not in accordance with empirical findings (see e.g., Frazier & Rayner, 1982; Meseguer et al., 2002), it is very likely that regressions are not merely a response but rather a solving mechanism to these problems (see also Schotter et al., 2014; Metzner et al., 2016). Thus, the eye movement itself has a function and the role of regressions in sentence interpretation is directly tied to this function.

Several reasons why a regression is launched have been discussed in the literature, e.g., difficulties in postlexical integration (Reichle et al., 2009), difficulties in word identification (Vitu & McConkie, 2000; Engbert et al., 2005) or difficulties in syntax / semantic processing (Frazier & Rayner, 1982). But crucially, none of these functions can account for all regressive eye movements occurring during sentence reading. However, because all regressions share the same characteristics (e.g., an eye movement against the intended reading direction, rereading of former sentence material, etc.), it should be possible to define one unifying principle underlying all regressive eye movements (with the exception of corrective saccades). In particular, this also satisfies von Ockham's principle of parsimony that constitutes good scientific practice.

According to the FC model, the function of a regression can be derived from the properties of the eye movement itself, which is to send the eye's fovea to a certain part earlier in the sentence, taking in additional visual input. Thus, the FC model proposes that the function of a regression is to gather additional information relevant in the course of sentence interpretation. In particular, since word identities are the crucial part of sentence interpretation, the FC model assumes that with a regression, the reader is aiming to get more visual information about the identity of words that have already been processed earlier, at least to some extent. This also means (although it is not explicitly stated) that regressions are not necessarily linked to processing difficulties, but may also be triggered by top-down influences such as tasks that require a deeper processing of the sentence. This stands in clear contrast to the accounts proposed in the literature

(Vitu & McConkie, 2000; Engbert et al., 2005; Frazier & Rayner, 1982; Reichle et al., 2009).

Since these assumptions of the FC model are perfectly in line with the working hypothesis proposed in the beginning of this thesis, and because they seem very plausible from a reading as well as language-oriented perspective, we agree with this assumption and propose within the *Information Gathering Framework* that the function of all regressive eye movements is to gather additional information about the identity of words[11].

3.1.3 The language model

The FC model proposes that because word identification is based on noisy visual information, *"word recognition may be best thought of as a process that is never ‚completed'"* (Bicknell & Levy, 2010, p. 1170). Accordingly, the reader uses the noisy visual input to compute the probability of the identity of a word on the basis of its language model (that is organized by frequency), but always a level of uncertainty remains.

Although we agree on the assumption of incomplete word recognition, we may ask whether noisy visual information is in fact the major determinant of word identification that may explain the variability in fixation durations. In particular, there exists convincing evidence that the decoding of visual information occurs very rapidly. That is, even when a word is presented for only 60 ms and masked afterwards, reading behavior is unaffected (Ishida & Ikeda, 1989). It seems therefore more plausible to assume that word identification is mainly affected by the retrieval of the lexical information (as also proposed by the SWIFT and E-Z Reader model) instead of problems with the visual perception. Thus, longer fixation durations would (primarily) correlate with an increase of lexical information and not with an increase of visual information, because the visual information is decoded rapidly.

This assumption, of course, requires a more complex language model than the simple bigram frequency model underlying the FC model that has limitations in the face

[11] Here we propose an account that is word-based as most models of eye movement control are (SWIFT, E-Z Reader etc.). But we think that this a simplification though because often single words can be better viewed in combination with other words, either as a fixed unit (as in chunks) or as a syntactical phrase that has a common function within a sentence (e.g. in noun phrases). This view is often held in sentence processing accounts that focus on phrases rather than words.

of the full complexity of linguistic processing during sentence reading as well (as discussed in chapter 2.2.3). Specifically, the lexical representation stored in the memory has to be viewed as an (theoretically) infinite bundle of features, containing information about the word's orthography, phonology, meaning, morpho-syntax as well as its constituent binding preferences (see Figure 2 for a schematic illustration; c.f. also Perfetti [2007] who introduced this idea as the concept of lexical quality in order to explain differences in language skill between individuals). Thus, whereas the FC model proposes that increased fixation times enable the reader to decrease noisy visual input, we make the claim within the *Information Gathering Framework* that increased fixation times primarily enable the reader to retrieve more information about a word's identity from the lexicon.

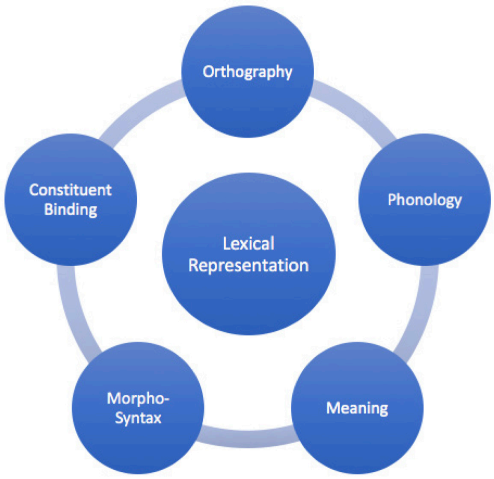

Figure 2: *A schematic illustration of the lexical representation of a word that is stored in the memory. This representation comprises the five linguistic levels orthography, phonology, meaning, morpho-syntax and constituent binding (see also Perfetti, 2007). Each of these levels consists of several (or infinite) more features. The lexical quality level describes the amount of information about a word that is currently retrieved from the lexicon.*

Whereas the term 'lexical representation' refers to the stored entry of a word in the lexicon, the term 'lexical quality level' describes the amount of information about a word that is currently retrieved from the lexicon. Typically, the amount of information (and thus the lexical quality level) continuously increases during a fixation, because a fixation allows for the retrieval of lexical information on the basis of the visual input[12].

[12] This does not mean that the whole time during a fixation is needed to encode the visual input (because this happens very fast, as mentioned earlier) but that the visual input allows for the retrieval of the lexical information.

However, once the eyes have moved to the next word, no additional information can be received[13] and the quality level is then continuously decreasing over time due to interference from other words and due to a decay[14] (Lewis & Vasishth, 2005; see Figure 3 for a schematic illustration). Also note that the lexical quality level of a word (as the confidence level) is never reaching the full quality level because the retrieval of the information from the lexical entry can by definition never be completed.

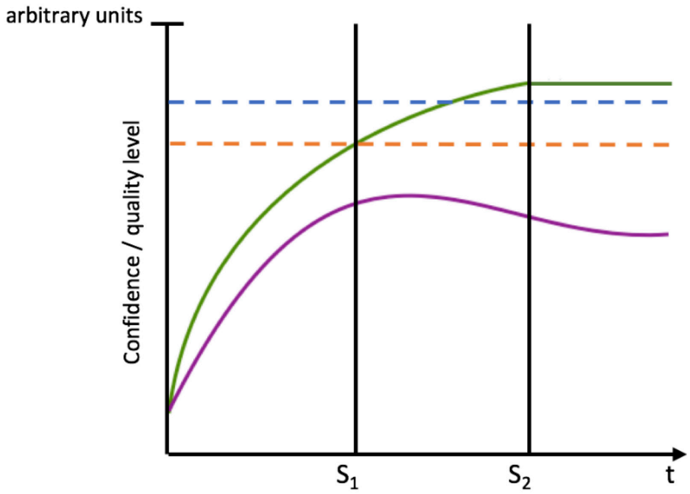

Figure 3: *Schematic illustration of the confidence / quality level of a single word during a typical sequence of two progressive saccades: Whereas the confidence level is continuously increasing and asymptotically approaching the full confidence level, the quality level decreases due to interference and decay after the eyes moved to the next word. Legend: purple = quality level, green = confidence level, orange = forward threshold, blue = backward threshold, S_1 = saccade to word n+1, S_2 = saccade to word n+2, t = time.*

It is obvious, however, that the lexical quality level depends on the lexical representation that is stored in the memory. Thus, the size and the specification of this lexicon[15] in particular, may affect the process of word identification in two ways. First, the

[13] This assumption might be problematic because there is evidence showing that the masking of the immediately preceding word does also affect reading behavior (e.g., Rayner, Well, & Pollatsek, 1980). This suggests that the visual information of word *n-1* is still used for word identification. However, it is unclear if this information affects the retrieval of lexical representation (as indicated by the lexical quality level) or processes that primarily require attention (as the computation of confidence levels) which are assumed to continue after the eyes have moved to the next word.

[14] It could also be that an eye movement within a word is performed (intra saccade). In this case, the quality level of a word is also increasing after the have eyes moved. But since we are focusing on inter-word saccades because intra-word saccades have to be viewed as functionally different from inter-word saccades (see chapter 1.1), we just aggregate all saccades within a word to a single inter-word saccade (thus talking in fact about first pass times, as most models of eye movement control in reading do).

[15] The size and specification of the lexicon do highly interact but are not the same. Thus, it could be possible that someone has a very small size of the lexicon in general (e.g., due to little

lexicon affects the amount and accuracy of information about a word's identity that can be retrieved from the visual input. Thus, a word representation that is stored in the lexicon may consist of a few features only if the word is not well known by the reader. Alternatively, it may consist of many features if the reader has a very detailed representation of a word (consider, e.g., the representation of technical terms). Also, the associated features could be correct or not (e.g., in phonology). Second, the lexicon affects the time it takes to retrieve this information from the lexicon because the word entries in the lexicon are sensitive to frequency (Forster & Chambers, 1973; Marslen-Wilson, 1990). That is, the more frequently an entry or its features are retrieved, the more easily (and swiftly) this information can be retrieved again.

In addition, the size and specification of the lexicon is also assumed to vary among individuals, depending on individual language exposure, language skill and the language under consideration. Thus, given a certain amount of time, the lexical quality level of a word is expected to vary amongst readers, as well as the time taken to reach a certain level of quality.

Note also that the lexical quality level is highly correlated to the level of confidence, but these two parameters are not the same. A poor reader could have a high confidence in a word's identity although it is ambiguous (e.g., in meaning). But due to a small lexicon which implies a representation of a few features only, the reader is not aware of these alternative interpretations. Accordingly, a proficient reader could have a low confidence into the same word's identity because he takes into account several potential ambiguities that the poor reader is not aware of. In addition, a highly predictive context may also effect that less information (and thus a lower lexical quality level) is needed to confirm this prediction and reach a certain level of confidence. This explains why fixations on highly predictive words are shorter than those on unpredictable words (see next section).

3.1.4 The computation of confidence levels

According to the FC model, the reader computes a confidence level of a particular word on the basis of its language model. This confidence level of a word is updated as more information in the course of the sentence is given. In particular, the FC model proposes

language experience) but that he has a very detailed representation of words that are in specific fields of which he is an expert in.

that additional information may also cause the confidence in a previous word's identity to fall. In response, a regressive saccade to this particular word is triggered, whenever its confidence level declines under a certain threshold.

As a consequence from the underlying bigram frequency model and the focus on reducing noisy visual input, the FC model does not view the computation of confidence levels as a matter of language processing. In particular, the FC model does not specify the lexical representation of a word in more detail, so that linguistic computations do not affect fixation durations.

This is unquestionably a very unrealistic assumption because reading is not only a visual decoding task but rather aims at the comprehension of a linguistic message. As already discussed above, the major determinant for fixation durations seems therefore to be linguistic processing instead of visual perception. In particular, because language processing is, like all cognitive processes, a time consuming task, decades of neuro- and psycholinguistic research have focused on determining the exact time course of language processing. This time course of language processing is, however, likely to correspond (at least partly) with the language hierarchy, with lower linguistic levels being processed earlier than higher linguistic levels.

Within the *Information Gathering Framework* it is therefore assumed that the computation of confidence levels (as with the computation of the lexical quality levels) is based on linguistic processing and takes a certain amount of time. During this time, the confidence level of a word typically increases (asymptotically approaching but never reaching the full confidence level), because more supporting evidence is given from the information of the lexical representation (see Figure 3 for a schematic illustration). According to the limited focus of attention which we will discuss in more detail in the next section, the computation of the confidence level of word *n* is restricted to the time the eyes fixate on word *n* and word *n+1*. After the eyes have moved to word *n+2*, no further computation of word *n*'s confidence level can take place and the confidence level remains stable over time.

Importantly, the *Information Gathering Framework* follows the "one-system hypothesis" of grammatical theories and language processing models (see Lewis & Phillips, 2015, for a recent discussion) claiming that the time course that is proposed within the framework corresponds to the time course of linguistic processing and in turn to the linguistic hierarchy. Thus, the earlier in time, the lower the linguistic level of the lexical representation, of which the computation of the confidence level is based on.

A challenging question, however, is how the confidence level is exactly computed. In the FC model, the confidence level is calculated on the basis of the noisy visual input by using a Bayesian inference term that computes the distribution over possible identities. Importantly, the FC model uses only bigram frequencies and no further linguistic information. Whereas this architecture of the FC model was chosen to allow for a formalized integration into a computational account, the situation within the *Information Gathering Framework* is much more complex because it explicitly claims that language processing is the major factor of eye movement control in reading. However, to map the complexity of language processing in full detail is beyond the scope of the present framework. Thus, the mechanisms proposed here are a simplification, but may be elaborated in more detail by future research.

For the current purpose, it is assumed that the confidence level is computed by matching the features of the lexical representation with the predictions of former sentence material on the basis of explicit production rules (Newell, 1973). These production rules represent all procedural knowledge (grammatical knowledge) and set condition–action pairs. For example, if an inanimate noun (e.g. *the table*) is encountered as the initial argument in an English sentence (*condition*), the production rules predict that a verb (*action*) will follow in the course of the sentence. More precisely, they predict that this verb should agree with the argument in number (singular), comes with an inanimate subject, and so on. If a verb like *talks* is encountered next, this leads to a violation of production rules because *talks* requires an animate subject. On the other hand, if a pronoun like the word *which* is following, it induces a relative clause. In this case, the production rules are not violated and the action (the expected verb) is simply postponed. Also, not every condition-action pair is mandatory; some pairs are just optional (e.g., the indirect object of verbs like *write*: *He writes a letter (to his father)*). If the evidence provided by the lexical representation matches the predictions made on the basis of the production rules, a high confidence level is computed. If the production rules are violated by contrast, it leads to a low confidence level. Accordingly, if the context is highly predictive, less lexical information and thus less time is needed to reach a certain level of confidence which results in shorter fixation durations.

Crucially, there are two different scenarios that trigger a regressive eye movement. In the first scenario, the computation reveals that the confidence level of word *n* falls under the so called forward threshold (that defines the level of confidence that is needed to proceed), although the eyes have already moved to word *n+1*. This happens if the predictions of the production rules are violated by the information from the lexical representation. In the second scenario, the computation causes the confidence level of word *n* to not reach the so called backward threshold (that prevents a regression from happening) within the fixation of word *n+1*. This happens if the optional predictions of the production rules are not met or other expected supporting evidence for the identity of a word is not given. We will explain these two scenarios in more detail in section 3.1.7, after we have outlined the assumptions about the limited focus of attention and the two control mechanisms.

3.1.5 Limited focus of attention

The FC model takes into account the limitations of the visual field in order to compute the degree of noisiness for the visual input, but it is not specified with regard to the focus of attention. However, because the underlying language model is restricted to bigram frequencies, the confidence level of a word can only be affected by the visual information about the subsequent word.

Within the *Information Gathering Framework*, the visual field also shapes the amount of visual information that is available to the reader during a fixation and that is used for the computation of the lexical quality level. But in addition, it is assumed that the computation of confidence levels always requires attention, so that not the confidence levels of all words in a sentence can be monitored in parallel.

In particular, recent research on the basis of SAT (speed accuracy trade-off) experiments has indicated that the focus of attention is very limited, covering only two chunks (McElree, 2006). We therefore assume within the *Information Gathering Framework* that the focus of attention is restricted to the word of the current fixation (W_6 in the example below), and the word before (W_5 in the example below) which means that only the lexical representations of these two words can be used in parallel to compute the confidence levels[16] (see Figure 4).

In particular, this implies that the confidence level of W_5 is computed on the basis of the lexical information only when the eyes fixate on W_5 and W_6, whereas the alignment with higher-order predictions involving syntactic and semantic processing primarily

[16] There is evidence that attention also covers W_7 when fixating W_6 (as indicated by the parafoveal-on-foveal effects) but because the quality level of the lexical representation of W_7 is only small (given the time interval it takes to compute lexical quality) we assume that word W_7 is only partially in the focus of the current attention.

happens after the eyes have moved from W_5 to W_6 because the time course of language processing (and thus the computation of confidence levels) is assumed to correspond to the language hierarchy, at least partly. This assumption of spill-over processing is shared by several current models of eye movement control like E-Z Reader and SWIFT, that adopt the claim (on the basis of clear empirical findings) that the processing of a single word is not restricted to the time upon which this word is fixated (in contrast to the very strong eye-mind hypothesis proposed by Just and Carpenter, 1980). However, after the eyes have moved to W_7, the lexical representation of W_5 is not in the focus of attention anymore, thus the confidence level cannot be computed on the basis of this information anymore and remains stable.

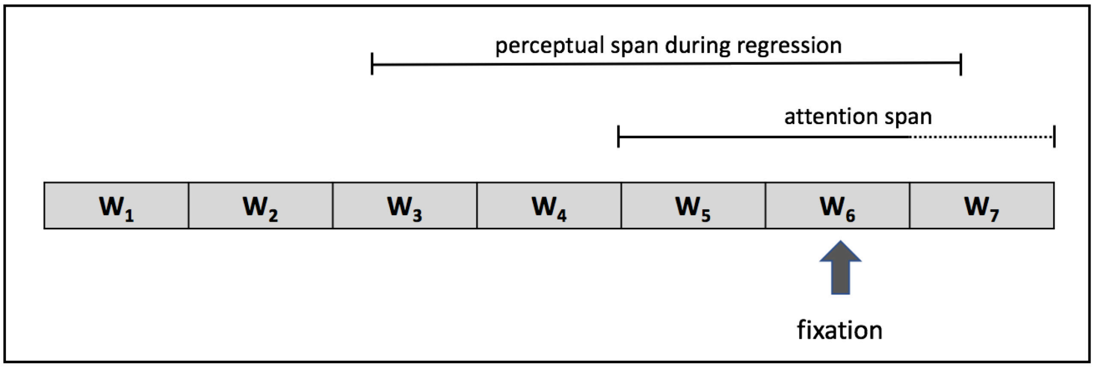

Figure 4: *Schematic illustration of the attention and perceptual span. Only word representations within the attention span can be used to compute confidence levels during the current fixation. If a regression is triggered on W_6, the attention shifts and the lexical representations of the word within the perceptual span (here W_3, W_4 and W_5, assuming 5 letter words) can be retrieved and used for a new computation of confidence levels. If this computation reveals no clear result, a regression target is selected on the basis of a strategy. Note that if W_1 or W_2 becomes the target of a regression, they are always assumed to be selected on the basis of a strategy because they are beyond the perceptual span. See text for further details.*

If a regression is triggered on W_6, however, because the confidence level of W_5 is not reaching the defined threshold (see section 3.1.7), an attention shift to the left is performed that allows for the retrieval of previous lexical representations and a re-computation of confidence levels. This assumption follows from research on the perceptual span in reading, which describes the area around the current fixation where disruptions of the text still affect reading speed.[17] The perceptual span has been examined by using

[17] Evidence from attention research suggests that there exists only a division between the current focus of attention (comprising only two words) and the remaining part of the sentence (see e.g., McElree, 2015, for an overview). In reading research, however, large evidence has strengthened the role of the perceptual span for attention as an additional parameter (for a review see Rayner, 2009). It is thus an open question how these two conflicting assumptions can be brought together. For the current framework we therefore propose a threefold division into

gaze-contingent moving window paradigms in which only a defined window around the current point of fixation is visible to the readers and the remaining parts of the sentence are masked with a row of Xs (see Rayner, 2014, for a recent review).

Several studies have shown that the perceptual span comprises 3 to 4 letter spaces to the left of the fixation (McConkie & Rayner, 1976; Rayner, Well, & Pollatsek, 1980) and 14 to 15 letter spaces to the right of the fixation during reading (McConkie & Rayner, 1975; Rayner & Bertera, 1979). Because the perceptual span is not a restriction of the visual system per se, but is rather affected by attentional processes (as for example indicated by the finding that systematically increasing the font size of the letters to the right or left of the fixation does not reduce the perceptual span: Miellet, O'Donnell, & Sereno, 2009), it has been hypothesized that the perceptual span changes when making a regressive eye movement. This hypothesis has been confirmed by research of Apel and colleagues (Apel, Henderson, & Ferreira, 2012), who showed that the size of the perceptual span switches toward the direction of the eye movement which also implies a shift of attention to the left. Although the authors did not answer the question of the actual size of the perceptual span to the left of a fixation during regressions (and of course more research is needed), we assume for our purposes that it encompasses 15 characters to the left, according to the size of the right perceptual span in progressive eye movements.

It follows that for the architecture of our *Information Gathering Framework*, when making a regression, the lexical representations of the words within 15 characters to the left of a regression can be used to re-compute the confidence levels and to guide the regression target selection (see section 3.1.8).

3.1.6 The confidence level is monitored by two independent control mechanisms

The FC model proposes that the generation of eye movements is monitored by a simple control policy that sets two different values of confidence that cause an action. If the first value is reached, a forward saccade to the next word of low confidence is initiated. If the confidence level falls under the second value, a regressive eye movement to this particular word is triggered.

---

the focus of attention (i.e., the words that are processed in parallel), the perceptual span (i.e., the words that can be covertly retrieved for re-computing of the confidence level) and the remaining part of the sentence. It will be a subject for future research to examine in more detail how the attention during reading is shifted and how attention is used to select the regression target.

In the *Information Gathering Framework* the actions are also controlled by two independent thresholds, which we refer to as the forward and the backward threshold respectively, which are set by the forward and backward mechanism respectively. Due to the restricted focus of attention, only the confidence levels of word *n* and word *n+1* are affected by these two thresholds.

The first (forward) mechanism defines the level of first pass confidence, namely the amount of evidence about word *n*'s identity that is retrieved in first pass reading and assessed to be sufficient for the current sentence interpretation. When a certain level of confidence is reached, the eyes move to the next word. Since the computation of confidence levels is closely correlated to the time course of language processing, it is assumed that the forward threshold should be more sensitive to low level (e.g. orthography) as opposed to higher level linguistic information (e.g. syntax or semantic processing). It should be noted, however, that the sensitivity for different linguistic levels depends on the actual adjustment of the forward threshold which is assumed to vary among tasks and individuals.

It is further proposed that this forward control mechanism works in a highly automatic manner, per default targeting the next word. This automatic saccade generation is canceled and the eyes move to word *n+2*, if parafoveal processing already reveals a certain level of confidence for word *n+1*. The forward control mechanism proposed here is compatible with current models of saccade control like SWIFT (Engbert et al., 2002; Engbert et al., 2005) that assume a) parallel processing of different words, b) largely automatic generation of progressive eye movements and c) word identification as the core function of saccades in reading.

This forward threshold in particular mediates between speed and accuracy: If the threshold is set down, the reading speed is increased but accuracy also suffers. If the threshold is set high, by contrast, the accuracy is higher but at the expense of reduced reading speed.

The second (backward) mechanism defines the level of uncertainty, namely the level of uncertainty about the previous word's identities that causes a regression. But in contrast to the forward control mechanism, this backward mechanism is highly linguistically controlled.

One task of the backward mechanism is to specify a general threshold for triggering a regression (similar to the forward mechanism). Thus, a regression is performed whenever the level of confidence for a word does not reach a certain threshold. Recall that the confidence levels are by default only computed for the words within the attention span, so that only the confidence levels of these words may cause a regression. But in addition, the backward mechanism further monitors the selection of regression targets by shifting the attention to the left and by re-computing the confidence levels of previous words, more precisely of words within the perceptual span. If the re-computing reveals that the confidence level of one word falls under the backward threshold, a regression to this word is performed. In the case the confidence level of more than one word or no words falls under the backward threshold, the regression target is selected by using experience based strategies (we will explain this procedure in more detail in section 3.1.8).

Because the backward threshold is temporally delayed, it requires a higher level of confidence than the forward threshold. Thus, the backward threshold is assumed to be sensitive to later and therefore higher-order linguistic processing, at least compared to the forward threshold. But again, the sensitivity depends on the actual adjustment of the backward threshold.

Although the forward and backward control mechanisms often interact, they are assumed to be independent and may be adjusted separately. Thus, there may exist a first pass strategy that allows for relatively superficial reading, but this does not necessarily mean that at the same time the probability for regressions increases. In addition, both control mechanisms are assumed to be sensitive to top-down influences like tasks that may reduce or increase the thresholds for first pass reading times and regressions. Bicknell & Levy (2010) for example showed that the most efficient reading strategy (i.e., the one that leads to highest comprehension accuracy) is one that allows for a lower level of confidence in first pass and increases the probability for regressions at the same time.

3.1.7 How a regressive eye movement is triggered

To illustrate the properties of these two control mechanisms in more detail, we will now first set out what exactly triggers a regressive eye movement within the *Information Gathering Framework.* Note that this procedure clearly differs from the one proposed by the FC model because it tries to also model the time course of language processing and the limited focus of attention. In particular, the FC model only knows one scenario that causes a regression, namely, if evidence from input later in the sentence causes low confidence of precious entities, whereas the *Information Gathering Framework* proposes two different regression scenarios.

First of all, recall that the focus of attention is restricted to word *n* and word *n+1*, which we will refer to as word 5 (W_5) and word 6 (W_6) respectively, according to Figure 4. Now, let us consider what happens when the eyes fixate on W_6.

First, the letter string of W_6 is perceived and the lexical quality level is computed. Because this takes a certain amount of time, the quality level (the amount of lexical information) of W_6 is typically increasing during the fixation on W_6. In addition, the confidence level is computed by matching the features of the lexical representation with the predictions according to the production rules. Thus, the confidence level is also typically increasing during a fixation. If it reaches the forward threshold, an eye movement to the next word (W_7) is performed.

Second, at the same time the quality level of W_5, that peaked right after the eyes moved to W_6, is now slowly decreasing due to interference and decay. The computation of the confidence level on the basis of this lexical representation, by contrast, stops and the confidence level remains stable. Note that the confidence of W_5 already reached the forward threshold because this triggered an eye movement to W_6. But the lexical representation is now used to match also higher-order predictions that need more time to be computed. Thus, the confidence level of W_5 is typically also increasing during a fixation on W_6 and reaches the backward threshold. If this backward threshold is reached before the confidence level of W_6 reaches the forward threshold, everything is fine and an eye movement to W_7 is performed (see Figure 5, Pattern 4).

There are two scenarios, however, where this forward walk through the sentence is disrupted. First, if the computation of the confidence level of W_5 reveals great difficulties that question the confidence into a word's identity so that the confidence level of W_5 even falls under the forward threshold. This causes a regression, independently of the confidence level of W_6.[18] We assume that this happens especially due to higher order parsing problems that do not allow for an integration of the current word into the adopted sentence structure.

Second, if the computation of the confidence level of W_5 takes more time because the predicted evidence is missing or more information has to be taken into account, the

[18] Note that also the computation of the confidence level of W_6 may detect very early integration difficulties (before the forward threshold is reached). In this case, either fixation duration on W_6 increases or the information is used for the computation of the confidence level of W_5 because both words are in the focus of attention. In this case, even information about W_6 may cause that the confidence level of W_5 falls under the forward threshold and a regression is triggered.

confidence level of W_6 may reach the forward threshold before the confidence level of W_5 reaches the backward threshold. This also triggers a regression. We assume that this happens especially at the end of a sentence where the sentence is evaluated as a whole (see chapter 3.1.3).

To illustrate the two scenarios, consider, for example, one of the stimulus sentences again used in the famous experiment by Frazier & Rayner (1982):

(3) Since Jay always jogs a mile and a half seems like a very short distance to him.

When the eyes fixate on the word "*seems*", the visual string is received, the lexical information is retrieved from the memory and the lexical quality level increases. Because the string is acceptable on lower linguistic levels (since it is a well-formed and highly frequent word of English, for example), the confidence level reaches the forward threshold soon and a progressive saccade to the next word "*like*" is performed. However, during the fixation of "*like*", the computation of the confidence levels of both words (*like* and *seems*) is performed in parallel. For the word "*seems*" the production rules now yield an error on the syntax level because the word cannot be integrated into the current sentence structure. As a response, the confidence level for the word "*seems*" decreases and falls under the forward threshold which causes a regressive eye movement, independently of the confidence level for the word "*like*" (scenario 1).

For the second scenario, consider a fixation on the last word of the sentence: "*him*". Here the confidence levels of "*him*" and "*to*" are computed in parallel and no violation of production rules is detected. However, because the eyes parafoveally receive the punctuation which signals the end of the sentence, the reader evaluates the meaning of the whole sentence in order to increase the confidence level of the word "*to*". This may lead to a delay because the reader might have expected more evidence that was not given in the course of the sentence (for example, more information about *Jay*). In the meantime, the confidence level for the word "*him*" has reached the forward threshold, although the confidence level for "*to*" has not reached the backward threshold. This also causes a regressive eye movement in response (scenario 2).

In both cases, however, a regressive eye movement is performed because the available information about current word identities is assessed to be insufficient for the current sentence interpretation. Thus, a regression always aims to gather additional evidence about a word's identity and to increase the level of lexical quality of previous words by making a shift of attention. This in turn normally also increases the confidence level of these words.

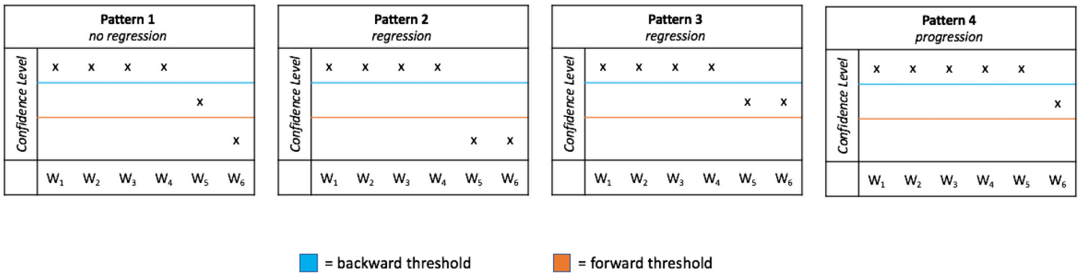

Figure 5: *Potential patterns of confidence levels. Each pattern represents the confidence levels of six words (W_1 to W_6) during a fixation on W_6. The blue line shows the backward threshold and the orange line the forward threshold.*

3.1.8 How the target of a regressive eye movement is selected

We have seen that a regressive eye movement is triggered on W_6 whenever the confidence level of W_5 does not reach the backward threshold within a certain time window or the confidence level even falls under the forward threshold. However, the crucial question is how the target of this regressive eye movement is selected.

The FC model predicts that the regression always targets the word with the confidence level under the backward threshold. But since the computation of confidence levels may only cause the confidence into the identity of the directly preceding word to fall (due to the bigram frequency model), regressions are always targeting word *n-1* (as is also assumed by E-Z Reader 10, for example). However, this assumption is just a simplified approximation and apparently, the predicted target pattern does not fit with the patterns reported in the literature (see chapter 2.4 for a discussion). Accordingly, the opportunity within the *Information Gathering Framework* to always regress to W_5 does not fit with empirical evidence although a majority of regressions are actually targeting W_5. In addition, the word in the sentence where problems become apparent does not always correspond to the word that causes difficulties. A very prominent example are garden path sentences (as discussed earlier), where difficulties are often caused by a misinterpretation of a word earlier in the sentence. In this case, a re-inspection of W_5 would not help to solve the problem and since we assume that the function of a regression is to *solve* the problem, this is not a plausible mechanism.

Another opportunity would be to select the word with the lowest quality level as the target for the regression because there is an increased likelihood that more evidence (provided by the lexical representation) about this word would help to increase confidence. However, there are also difficulties with this assumption: As already discussed, the quality level and the confidence level are not the same. Thus, a low quality level does not automatically cause a low confidence level. In addition, this assumption would lead to the conclusion that words earlier in the sentence / text are more likely to become the target of a regression because the quality level is low (due to the decrease over time). This prediction, however, is not supported by the empirical findings either.

Thus, we assume that in the case of a regression a re-computation of the confidence levels of previous words takes place and that the selection of a regression target is linguistically constrained. In particular, it is assumed that if a regression is triggered, the attention is shifted to the left of the fixation and the confidence levels of the words within the perceptual span (14-15 characters) are computed again, on the basis of their lexical representation and the applied production rules. If this causes the confidence level of a particular word (in our example of Figure 5 the confidence level of W_3, W_4 or W_5, given 5 letter words) to fall under the backward threshold, this word is selected as the regression target.

In the case the confidence level of more than one word falls under the backward threshold, the target (W_3, W_4 or W_5) is selected by the backward control mechanism on the basis of experience based strategies. The same happens if none of the words' confidence levels within the perceptional span falls under the backward threshold. In this case, the confidence levels of all words in a sentence fall under the backward threshold (as some kind of 'chaos response' because the cause of the problem cannot be determined by the reader) and the backward control mechanism selects the regression target on the basis of a limited set of strategies. Note that it is likely that a target selection based on strategy is more the rule than an exception.

The limited set of selection strategies is based on language experience and aims to define the most efficient way to gather the required information, without taking into account the details of the lexical representation itself. *Most efficient* is defined as the combination of speed and accuracy, which means that the strategy is the fastest way to find the most relevant information in the absence of explicit knowledge, taking into account the speed-accuracy tradeoff. *Limited set* means that only a restricted number of strategies (maybe 3-5) and not a full variety of strategies exist. *Language experience* means that this strategy has been applied most frequently in the past and yielded good results, so that the reader when he is faced with a certain category of tasks, assesses the likelihood where the relevant information can be found on the basis of his language experience. *Strategy* means that the same type of eye movement (B) is performed when faced with the same task (A) – at least for a single reader – resulting in the simple condition term: *if A, then B*.

3.1.9 Explaining landing site patterns of regressions

If the regression generation mechanism works in the manner just described, then the eyes are sent to a specific location in the sentence that is selected by a linguistic control. Thus, they are not subject to a random or default strategy as proposed by the *time out hypothesis* of Mitchell et al. (2008). Furthermore, this target location can either be determined by the lexicon / production rules and thus by explicit linguistic knowledge (in the case a single target location can be computed) or in addition by a strategy based control mechanism.

In many cases, however, the landing site patterns of regressions reported in the literature (e.g., Frazier & Rayner, 1982; Mitchell et al., 2008; Meseguer et al., 2002; von der Malsburg & Vasishth, 2011, 2013) show a distribution that challenges the assumptions of such a strong linguistic guidance. But whereas factors other than linguistic properties have rarely been discussed in the context of regression landing sites, we think that it is necessary and worthy to take into account more aspects that may shape the landing site distribution, although linguistic computations are assumed to be the main determinant.

In the following we will discuss several additional aspects of which we think may influence the location where the eyes actually land when making a regression. However, very little research has been done in this field so far, so that the interaction between these factors is largely unknown. Thus, the assumptions in this section have to be viewed as preliminary considerations and are not integrated into the *Information Gathering Framework* yet. Nonetheless, we will try to provide some ideas as to how this may be accomplished and how the proposed framework may account for such factors. However, the major goal in this section is to generate a tool for further investigations in order to understand the landing site patterns of regressions in more detail.

3.1.9.1 Differences between individuals

Since experiments in cognitive psychology traditionally focus on insights into the human reading / language processing mechanisms in general, most statistical methods average about the results of single subjects and draw conclusions from the group level only (with some exceptions, of course). But the precise differences between individuals could also be systematic and provide insights into the human reading / language processing system. Fixation durations of skilled readers, for example, have often been reported to vary systematically among individuals (Henderson & Luke, 2014; Rayner, Li, Williams, Cave, & Well, 2007). In addition, von der Malsburg and Vasishth (2011; 2013) reported that many individuals had a preference for one of the rereading patterns. Although the authors do not discuss the reason for this preference, it appears obvious that these preferences may also be caused by systematic differences between individuals.

As we have seen, the *Information Gathering Framework* proposes that the selection of a regression target is linguistically controlled and not subject to a random mechanism. In particular, it is asserted that the target position is either selected on the basis of explicit linguistic knowledge or on the basis of experience based strategies. In addition, the computation of target positions requires attention and a retrieval of lexical information. Both factors, however, are known to vary between individuals. Thus, it may be that different regression landing site patterns simply reflect differences in linguistic knowledge and working memory capacities between participants, rather than other systematic linguistic processing strategies. We will now discuss these factors in turn.

Linguistic knowledge

Although most language processing models assume (at least implicitly) that all healthy people are competent speakers of their mother tongue, it has been found that not all of them are able to understand complex syntactic structures equally. Object relative clauses in English, for example, have often been found to be very hard to understand, even for English native speakers, because they make high demands on the language processing system (see, for example, Rayner, Carlson, & Frazier, 1983; King & Just, 1991). Importantly, these sentences are plausible, grammatical structures in English. Nonetheless, some readers are obviously not aware of the ambiguities and fail to find a grammatical interpretation of the sentence structure, often even after rereading the sentence several times. Accordingly, it has been proposed that linguistic knowledge might be an important factor for language processing in general (see, e.g., Wells, Christiansen, Race, Acheson, & Macdonald, 2009; Rayner et al., 2016) and for the interpretation of these sentences in particular, which has been also indicated by empirical evidence (MacDonald & Christiansen, 2002; Wells et al., 2009).

The differences in the ability to understand complex syntactic structures may also explain the results of Schotter et al. (2014; see chapter 2.3) who found that triggering a regressive eye movement did not necessarily lead to better comprehension: It is possible here that some readers were simply not able to diagnose the error correctly and perform a useful regressive eye movement. In particular, this points to the important role of linguistic knowledge for the selection of regression targets.

If we assume that regressions are an attempt to try and *solve* a problem, we must first acknowledge that in order to *solve* a problem, it is necessary to *diagnose* the source of the problem and direct the eyes to the position in the sentence *causing* the problem. And this, of course, requires a high amount of linguistic knowledge.

As we have already discussed in chapter 2.2.1, the sentence location where difficulties become apparent is not necessarily consistent with the location causing the difficulties. A very prominent example for this phenomenon are garden path sentences. Consider the following example (an adapted version of the sentences used by Rayner, Carlson, Frazier, 1983):

(7) The teacher sent the flowers that looked beautiful was very pleased.

In this object relative sentence, the assumed preferred interpretation according to the minimal attachment principle of the garden path model is the main clause reading that analyzes the verbal phrase *'sent the flowers'* as a predicate to the initial noun phrase '*The teacher'*. At the second verbal phrase *'was very pleased'*, however, this main clause interpretation turns out to be wrong and a reanalysis is performed (towards an object relative clause reading), which leads to increased processing costs at the point of disambiguation[19].

But critically, the disambiguation region is the location in the sentence where the problem can be *detected* but not where it can be *solved* (at least in the case additional information is needed and the problem cannot be solved covertly, i.e., without triggering a regression; see Fodor & Inoue, 1994, for a discussion of detection and reanalysis

[19] As becomes apparent from this example, the focus on single word identification frequently falls short because not every word carries an autonomous meaning. Rather, words appear in phrases and hierarchical dependencies which is why word identification better has to be viewed as an interactive process that involves the identification of several words instead of a single word. Another opportunity is to view the interaction between the identification of single words in terms of predictions, thus facilitating the identification of words in highly predictive contexts of other words.

processes during sentence reading as well as Friederici, Hahne, & Saddy, 2002, for a discussion on this issue within the ERP literature). In order to solve the problem, however, the sentence structure has to be reanalyzed (according to the garden path model) by interpreting the ambiguous verb phrase *'sent the flowers'* as the sub-clause verb phrase, thus leading to a reduced object relative clause analysis.

The question whether and how a reanalysis is performed during sentence interpretation has received considerable attention in the last 35 years of psycholinguistic research (see Fodor & Ferreira, 1989, for an overview and von der Malsburg & Vasishth, 2013, for a recent discussion of sentence processing models in the context of regression landing sites), and it is well beyond the scope of this thesis to discuss the different approaches in more detail. The precise description of this reanalysis process would also depend on our own underlying lexicon and the production rules which are currently not specified with regard to these aspects. But despite the diversity in the models of sentence processing, all of the approaches conclude (at least implicitly) that a high amount of linguistic knowledge is necessary in order to reanalyze a sentence efficiently.

Frazier and Rayner (1982; see also chapter 2.4.1) discussed this issue of reanalysis by investigating the landing sites of regressions. They used regressive eye movements as a tool to provide insights into reanalysis processes during sentence reading and implicitly assumed (according to the garden path model) that there is one common strategy shared by all readers. Interestingly, their results showed clearly that the majority of regressions targeted the ambiguous region, but that the results also revealed a high variability in landing sites. Further evidence for a variability in landing sites is provided by the findings of von der Malsburg and Vasishth (2011; 2013), who also reported different regression patterns that additionally varied among individuals (see chapter 2.4.3).

This points to the opportunity (although it has not been discussed or examined by the authors themselves) that regression patterns may vary as a function of language knowledge and the ability of diagnosing the potential error. Specifically, this means that the ability to diagnose the potential source of the problem requires a high amount of linguistic knowledge which is not available to all readers.

In terms of the *Information Gathering Framework*, the ability to solve a problem is closely linked to the production rules (and to a lower extent to the specification of the underlying lexicon) because such production rules allow for the integration of constituents into the sentence structure. As we have discussed above, the production rules define condition-action pairs. Thus, in the case where predictions are not met by the current input, the production rules may define the exact condition within the perceptual span that is not met by the action. This may be an open syntactic node, for example, that is not filled by an argument. The production rules therefore reduce the confidence in the condition-argument, because the action-argument is not provided and subsequently a regression to the condition-argument is performed. This regression aims to gather additional information about the condition-argument in order to increase the confidence level.

In the case where the appropriate production rules are not available or the wrong production rules are assessed (i.e., linking the missing action to the wrong condition), this may lead to two scenarios: First, a low confidence level for the wrong word is computed, or second, that the confidence levels of more than one word or even all words are reduced as some kind of chaos response, because the error source cannot be determined. According to the *Information Gathering Framework*, therefore, the backward mechanism selects the target position on the basis of a strategy that has been developed drawing upon the language experience.

Memory capacities

Besides the specification of the underlying lexicon and the production rules, the execution of regressive eye movements is also very likely to be affected by limitations of the short-term memory capacity. Recall the resolution of ambiguities in garden path sentences: The parser has to detect and to diagnose the error correctly based on his linguistic knowledge. But in order to fulfill this task, the reader needs "*a mental record of the relevant points of parsing ambiguity that can guide regressive eye movements*" (Meseguer et al., 2002, p. 552).

Memory capacities have often been proposed to affect language processing (see, e.g., Baddeley, 2003; Van Dyke & Johns, 2012; Gibson, 1998). King and Just (1991), for example, examined the processing of object relative sentences in English like (8) with readers of different working memory capacities (indicated by a high or low score respectively in a reading span test).

(8) The reporter that the senator attacked admitted the error.

The authors argue that there are three particular demands for the working memory capacity which should lead to increased processing efforts for readers with a low working memory capacity: First, the main clause is interrupted by the embedded clause, second, the thematic roles for the noun phrases have to be assigned properly, and third, two different thematic roles have to be assigned to a single syntactic constituent. The results suggest that low span readers had indeed more difficulties in processing object relative sentences compared to high span readers, which the authors interpret as evidence for their hypothesis.

Although the concept of the reading span is problematic (Waters & Caplan, 1996; Friedman & Miyake, 2004; 2005), differences in sentence comprehension between subjects of varying memory capacity have also been reported by authors using other measurements (e.g., Waters, 1996; Fedorenko, Gibson, & Rohde, 2006).

Besides sentence processing per se, memory demands have also been discussed in the context of rereading strategies. As already discussed above, the most efficient method for rereading would be to specify exactly the location in the sentence that provides the relevant information for the task at hand. But in the case that this single target selection is not possible (which probably happens frequently), a strategy might be a useful compensation mechanism. These strategies, however, differ with regard to their efficiency (recall: efficiency = speed and accuracy) but also with regard to their required knowledge. Lewis (1998; see also von der Malsburg & Vasishth, 2011), for example, discusses several reanalysis patterns (see also Frazier & Rayner, 1982) and points out that

> *…the backtracking strategy that places the least demands on memory is forward reanalysis by overt recomprehension from the beginning of the sentence. This strategy requires no memory for input or prior parsing states – it needs just enough memory for the parser to remember not to continue the same path. The drawback of this method is time, but it is a reliable strategy when all else fails. (p. 253)*

This fits well with the results of von der Malsburg and Vasishth (2011; 2013) who found that forward reanalysis is a very frequent reanalysis strategy in sentence reading. These findings also accord with the *Information Gathering Framework*, which proposes a limited set of efficient rereading strategies that are applied in the instance where a single target location cannot be computed.

In the *Information Gathering Framework*, however, memory demands affect not only the development of efficient compensation strategies but also the triggering of regressions themselves. Thus, the availability of lexical representations to the reader (as indicated by the lexical quality level) is thought to be a prerequisite for solving processing difficulties. In particular, recall the role of the perceptual span for the selection of regression targets. Only words within the perceptual span are assumed to be available for the re-computing of confidence levels. Importantly, the perceptual span is not restricted by visual properties but by attention and thus interacts with working memory. There is evidence, however, that the size of the perceptual span differs among individuals, depending on linguistic knowledge and memory capacities (see Veldre & Andrews, 2014, for a review). It is thus very likely that a reduced perceptual span also leads to differences in the regression behavior between individuals, with an increased probability for using experience based strategies because the error source cannot be determined.[20]

Further support for this view comes from the assumption that the computation of the sentence meaning itself is a highly demanding task, so that a low capacity reader has less resources to perform additional tasks such as computing alternative interpretations or maintaining the whole sentence structure (as proposed by King and Just, 1991).

To sum up, there are several reasons why it is plausible to assume that memory demands affect sentence processing and in addition, the execution of regressive eye movements. This is because eye movements are highly linguistically guided. Furthermore, the crucial role of the perceptual span for regression target selection within the *Information Gathering Framework* points to the possibility that individual memory capacities may affect regression landing sites.

Differences in the oculomotor system

A third possibility could be that the execution of eye movements by itself, i.e. the oculomotor system, is subject to individual differences (see next section). Rayner and colleagues (Rayner et al., 2016), however, argue in their discussion of speed reading techniques that there are clear physical limitations of the oculomotor system, so that the oculomotor system itself provides little room for individual differences (healthy subjects provided). The authors rather claim that *"[f]aulty language processing generally causes problems in eye movement programming, not the other way around*" (p. 9). Later, they

[20] Note again that the assumptions about the crucial role for the regression target selection conflict with the dichotomy of attention proposed by McElree and colleagues (see, e.g., McElree, 2015). We think, however, that this is justified by the empirical evidence about the variability of the perceptual span (see also footnote 17).

argue in the context of speed reading: "*The kind of practice that will help reading is practice that helps people to identify words and comprehend better, not just take in visual information faster*." (p. 28)

Thus, we may conclude that individual differences in linguistic knowledge and memory capacities have a great impact on the control of regressive eye movements, because the execution and target selection of regressions is strongly guided by linguistic factors. Individual properties of the oculomotor system by contrast, should play a minor role only (or even no role) in determining the landing position of regressions.

3.1.9.2 General factors that may affect regression targets

In the last chapter we have seen how differences between individuals with regard to linguistic knowledge and memory capacities may affect the selection of an appropriate regression target. But despite these individual differences, there may also be factors that affect the regression target process with only little variance between readers. Especially, whereas we have focused on the *selection* of regression targets so far, it is also possible that an intended target is not reached. In the following, we will discuss three factors that may cause such deviations from the intended target position, namely spatial memory, oculomotor error and visual salience. It is of course difficult to distinguish empirically between the intended target position and the actual landing position. Thus, more research is needed in this field using novel experimental designs.

Spatial Memory

We have already seen that linguistic processing places high demands on memory capacities. But linguistic processing beside, the execution of regressive eye movements itself also requires some spatial memory, i.e., the reader has to know at which location in the sentence the selected target word can be found.

In a series of experiments, Inhoff and Weger (Inhoff & Weger, 2005; Weger & Inhoff, 2007) examined the precision of task-induced regressions in reading single sentences. The readers were asked to respond to a question where the required answer was given at a specific location in the sentence they had previously read (and which was no longer visible – at least in some conditions). An example is given in (9):

(9) *My father is younger than my mother.*
Who was born earlier?

The authors showed that long-range regressions can be directed on the basis of stored linguistic and spatial representations, but that the accuracy of the eye movement declines depending on the visual distance. Thus, sufficient memory capacities may be likely to affect the precise location of regression targets.

The factor of spatial memory is not directly incorporated in the *Information Gathering Framework* as yet. It could be that the word representations also encode information about their position in the sentence, at least about their relative order. Inhoff and Weger refer to this idea as linguistic memory, i.e., that information about the relative order of words is stored but not their absolute spatial position (which they refer to as spatial memory).

Lewis (2005), however, proposes that there is no explicit serial order representation and that *"word-order constraints are deeply embedded in the structure of all the production rules, because they depend on a distinction between the word being processed now and what has come before (what must be retrieved)"* (p. 388). The question then is, how an eye movement can be performed to a precise location in the sentence if information about this location is not stored? This is particularly relevant for words outside of the perceptual span, where no visual information can be used to guide the target selection of the regressive eye movement.

We think that this is a critical question for modelling regressive eye movements (see also for example Eichenbaum, 2013, for a recent discussion of temporal and visuospatial aspects of memory in the brain) but we must leave it for future research to determine whether, and how spatial information is stored and used to guide regressive saccades.

Oculomotor Error

Another factor that may explain why – beyond the lack of spatial memory – an intended target position is not reached by the eyes, is the error of the oculomotor system (McConkie, Kerr, Reddix, & Zola, 1988). This is when the eyes overshoot or undershoot the length of a saccade, which can happen randomly as by all motor tasks or systematically. In general, after such an error a corrective saccade is performed in response. Accordingly, most of the current models of eye movement control in reading assume that oculomotor error affects the landing position of regressions. As a consequence, they have incorporated a term that accounts for such instances (see, e.g., the SWIFT model: Engbert et al., 2005).

Note, however, that the deviations from the intended target position are very small (i.e., normally within a word) and that only very few saccades actually land on a word that is not the target word. Thus, the impact of the oculomotor error is rather minor and does not influence the overall shape of the regression landing site patterns.

Visual Salience

As we have discussed above, the window of text where the eyes may gather visual information is limited physically by the properties of the oculomotor system and attentionally by the amount of information that can be processed at a time.

In the case the target position is within the perceptual span, this points to the question as to how far the eyes are guided by visual constraints of the target words. Friston and colleagues (Friston, Adams, Perrinet, & Breakspear, 2012) proposed the idea that saccades in general can be viewed as experiments: They try to test hypotheses about the visual input. In particular, it is assumed that these hypotheses are generated by a cognitive control, but that this cognitive control selects the target on the basis of properties of the visual input itself. Thus, it would be possible for a target word of a regression to be selected on the basis of its visual properties. Note also that Mitchell and colleagues (2008) found evidence that the layout of the text did influence the selection of regression targets.

The idea that visual information plays a role for the landing position of regression targets is also (at least partially) already incorporated into the *Information Gathering Framework,* because the quality of a word's representation also contains information about a word's orthography (i.e., its length and other visual features like upper- and lower case). When a particular target word has been chosen, either by the lexicon or by the backward control mechanism, the stored visual features could help the eyes to find this particular word in the sentence if the stored spatial information is not sufficient. In addition, certain visual features (e.g., font color) could make a word more salient in first pass reading, thus increasing the quality of a word's representation or assigning orthographic retrieval cues.

Another possibility could be that the backward control mechanism itself is sensitive to visual information, and that a target is selected by a strategy that relies on visual features. This could be useful in languages such as German, where the visual form of upper case letters often maps on a linguistic function, namely the head of a noun phrase. Accordingly, the backward control could have developed a strategy that selects words with certain visual features like upper case or uses the punctuation as a visual boundary (c.f. Hirotani et al., 2006) . However, there are at least two problems with this view: First, the reader needs a very precise spatial representation in the case the word is beyond the perceptual span. Second, it is questionable whether a strategy that targets on upper case is very effective, given the high variability of linguistic tasks.

Therefore, we have to leave it to further research to investigate whether and how visual features determine the target positions of regressive eye movements, and how these factors can be incorporated in the current model.

3.2 Applying the *Information Gathering Framework* to the findings in the literature

Having described the main properties of the *Information Gathering Framework*, it is now important to examine how the model can account for empirical findings of regressions that are reported in the literature. If the framework fails to provide a plausible explanation for the major results from eye movement experiments, it loses its justification even though the strength of a model cannot solely be assessed by the number of findings it covers (Seidenberg, 2012). Nonetheless, in the following we will discuss how the architecture of the *Information Gathering Framework* may account for a variety of critical empirical findings reported in the context of regressive eye movements.

3.2.1 Regressions to the immediately preceding word

Although the landing positions of regressions spread over the whole sentence, the majority of regressive eye movements target the word immediately preceding the currently fixated word (see e.g., Vitu & McConkie, 2000; von der Malsburg & Vasishth, 2011, 2013 for corresponding evidence). In particular, all current models of eye movement control discussed above (E-Z Reader 10, SWIFT, Model of falling confidence) only account for these instances.

Mitchell et al. (2008) argue (in favor of an automatic regression mechanism) that a regression from word *n+1* to word *n* is the "*smallest possible regression*". And of course, a regression to word *n* has some important advantages compared to target words that are farther away from the current fixation: First, the saccade is short and fast, so that less effort for its execution and control is necessary. Second, the target word can be processed parafoveally so that the saccade can be guided by using visual input. Third, memory demands are low because the word has been encountered immediately before. Also, in an automatic, default-like regression behavior, memory demands are almost not required.

In the *Information Gathering Framework*, however, we argue that regressions to the immediately preceding word can be explained more plausibly by a regression mechanism that is controlled by linguistic factors.

As already discussed in previous chapters, the *Information Gathering Framework* assumes that the computation of the confidence level continues after the eyes have moved to word *n+1* because the retrieval and integration of linguistic information takes time – an assumption that is in accordance with reading models like E-Z Reader and SWIFT (c.f. also the so-called spill-over effects: Rayner & Duffy, 1986; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). In particular, because language processing is hierarchically organized and this hierarchy is assumed to correspond to the time course of sentence interpretation, the computation of the confidence level of word *n* on word *n+1* encompasses primarily higher-order linguistic processing like lexical integration. Thus, an integration failure of word *n* will often become apparent only on word *n+1*.

If this integration fails because the predictions based on the production rules are not met, a regression is triggered. If the production rules reveal that more information about word *n* is needed (which is of course within the perceptual span and thus subject to linguistically based re-computation of confidence levels), the confidence level of word *n* is set down and a regression is performed (because the eyes already have moved to word *n+1*), targeting word *n* (see also Reichle et al., 2009).

Because there are many more instances in which the integration of word *n* fails due to wrong / less specified assumptions about its identity than instances where the integration fails due to wrong / less specified identities of previous words (which is the case, e.g., in most garden path sentences), the eyes very frequently regress to word *n.* This explains why the majority of regressions target the immediately preceding word.

In addition, the backward control mechanism could also have developed a strategy that selects the preceding word. Recall that the strategies applied by the backward control mechanism are assumed to be based on general language knowledge / experience and hence operate on frequency. Thus, in the case a low confidence level of more than one word (or no words) is computed, the backward control mechanism might select the preceding word, because this word often provides the most useful information in order to solve the processing problem.

This view is further supported by the findings of von der Malsburg and Vasishth (2013) that indicate that low-capacity readers were less likely to reread the sentences when faced with garden path sentences. Instead, they used rapid regressions to the word in the pre-disambiguating region more frequently[21]. Since these rapid regressions provide some advantages with regard to memory capacities (as discussed above), this strategy suits readers with low memory capacities.

---

[21] The results also indicate that most readers were not able to comprehend the sentences, thus it is very likely that the rereading pattern requires more linguistic knowledge / memory ca-

3.2.2 Properties of the target word

In their analysis of an eye tracking corpus of four adult readers, Vitu and Mc Comkie (2000) reported two relevant findings with regard to regressive eye movements: First, if a word has been skipped, the probability increases that this word becomes a target of a regression. Second, if a word is long and low in frequency, the probability also increases that this word becomes a target of a regression. In chapter 2.1 we have already discussed how these findings fit in general with an account that proposes that regressions are triggered in order to gather additional information. Notwithstanding this, we will discuss in more detail how the *Information Gathering Framework* may account for these results.

If a word is processed only parafoveally, the amount of information that can be retrieved from the visual input is reduced. The same applies for words that are very long or low in frequency which also delays the retrieval of word information (Forster & Chambers, 1973; Weekes, 1997). In the current framework, this reduced information leads to a lower level of lexical quality, albeit that the confidence level reaches the forward threshold. Thus, if input later in the sentence reveals some difficulties that cause the confidence into the previous words' identities to fall, then this happens especially for words with a low quality level because the backward mechanism determines which quality is not sufficient in order to fulfill the task at hand or to solve the problem[22]. In response, a regressive eye movement to this word of low quality is performed in order to increase the quality level which in turn increases the confidence level of this particular word.

In those rare cases of regressions where the increased quality level does not lead to a higher confidence level (e.g., in anomalous sentences), this increasing quality is used to re-compute the confidence level of the other words in the sentence and to trigger another regressive or progressive eye movement (or to finally abandon the attempt if no coherent sentence interpretation can be found).

Thus, the *Information Gathering Framework* predicts that words with a lower level of lexical quality are more likely to become the target of a regression which is in the line with the findings reported in the literature.

---

pacities than the rapid regressions. However, because the authors did no examine the relationship of regression patterns to comprehension accuracy (see chapter 2.4.3), this issue has to be left for future research.

[22] Recall, however, that a low lexical quality level does not necessarily lead to a low confidence level.

3.2.3 Sentence wrap-up effects

A clear deficit of eye movement models like SWIFT and E-Z Reader is that they only attribute regressive eye movements to processing difficulties. Whereas this of course covers a wide range of regressions reported in the literature, it ignores some important results at the same time.

Several studies provide clear evidence for an increased probability to regress from the end of a sentence (Frazier & Rayner, 1982; Meseguer et al., 2002; von der Malsburg & Vasishth, 2011; 2013). These so called 'sentence wrap-up effects' occur largely unaffected by sentence processing difficulties, or at least not showing up at the location in the sentence where difficulties are expected to become apparent (although other reading measures indicate difficulties at these locations, e.g., increased first pass reading times). Thus, these regressions cannot be directly attributed to failures of the lexical integration, for example. In addition, although the target positions of these regressions are rarely reported, most of these regressions seem to target the beginning of the sentence, thus resulting in rereading from the beginning (von der Malsburg & Vasishth, 2011; 2013).

As we have discussed in previous chapters, these findings are generally in accord with a framework that assumes that regressions are triggered in order to increase confidence into previous words' identities, since the reason why confidence has to be increased is not coupled to difficulties of integration per se. In particular, the *Information Gathering Framework* posits that regressions are triggered whenever the predictions made by previous input are not matched. This could either be that the current input conflicts with the predictions (which would lead to a decrease of confidence) or that expected evidence is missing (which would lead to a slower increase of confidence).

In the case of regressions from the final region we assume that the latter scenario takes place. Thus, if the eyes move to the final (or pre-final) word, the confidence level of this word is computed by matching the predictions. But in addition, the punctuation is also received from the visual input (at least parafoveally), which signals a sentence boundary. Sentence boundaries indicate that no additional input for the current sentence interpretation can be received and subsequently no prediction (condition-action pair) can be postponed to later input. Thus, at the end of a sentence an evaluation of the whole sentence interpretation takes place (Just & Carpenter, 1980; Rayner et al., 2000; Hirotani et al., 2006). In the case that this evaluation reveals that more evidence is needed in order to develop a coherent sentence interpretation, a regression is performed to compensate for this information deficit. Of course, the degree of evidence (and of confidence, respectively) into a sentence structure that is assessed to be sufficient (the backward threshold) may depend on factors like task or time pressure.

Since an evaluation of the whole sentence takes place without dealing with a concrete integration problem, it is reasonable to assume that not a single target position based on the production rules can be defined. In contrast, the regression strategy applied selects a target position on the basis of language experience. This prediction fits well with the regression patterns reported by von der Malsburg and Vasishth (2011; 2013), which show a clear tendency to regress to the beginning of the sentence and to read the whole sentence again.

3.2.4 Fixation times and regressions

At the beginning of this thesis, we discussed the functional differences between regressions and increased fixation times which are both considered to reflect comprehension difficulties. We argued that during a fixation the reader has to deal with the currently available information, whereas a saccade allows for the intake of additional input. This difference, we concluded, makes it plausible that these two mechanisms do not just sum up each other, but that they rather serve for different purposes in reading.

This distinction is further supported by the counterintuitive finding that fixations before regressions tend to be shorter, which was first reported by Altmann and colleagues (Altmann, Garnham, & Dennis, 1992). According to the assumption that both measures just reflect processing difficulties, there is no reasonable explanation for such a behavior. By contrast, the *Information Gathering Framework* provides some ideas how these two measures can be distinguished. This distinction is especially accomplished by assuming that the forward and backward threshold are sensitive to the processing of different linguistic levels, due to the fact that they mirror the time course of linguistic processing.

Within the *Information Gathering Framework* it is assumed that the quality level of word n increases only during the fixation of word n. After the eyes have moved to word $n+1$, no additional information can be received and accordingly the lexical quality level cannot increase any more. Note also that the retrieval of information from the visual input and the retrieval and computation of lexical representations is a time consuming process, which explains why the lexical quality level is increasing during a fixation and not just a fixed task that is done or not (see chapter 3.1.2).[23]

In the current framework in particular, a progressive saccade is triggered whenever the confidence level reaches the forward threshold. This forward threshold defines the amount of information about a word that is assessed to be sufficient for the current sentence interpretation. If this information is assessed to be not sufficient, the fixation time is increased. Crucially, the forward threshold does not require the integration of all linguistic levels, but is rather reached once lower levels of linguistic processing are completed (recall that a correspondence between time and linguistic hierarchies is assumed).

A regression, by contrast, is triggered whenever the confidence level of word *n* does not reach the backward threshold during a fixation on word *n+1*. Thus, because the computation of the confidence level continues after the eyes have moved to the next word (in contrast to the quality level), the backward threshold is sensitive to higher-order linguistic processing. Accordingly, a regression is not triggered if lower linguistic levels violate the predictions of the production rules, because in this case the confidence level would not have reached the first pass threshold. Thus, increased fixation durations in first pass reading and an increased probability to regress should reflect a response to different processing difficulties.

Let us now briefly discuss the finding that fixations before regressions tend to be shorter. As we have seen, the triggering of a regression is also a matter of time: A regression is triggered either when the confidence level of word *n* does not reach the backward threshold before the confidence level of word *n+1* reaches the forward threshold, or when the confidence level falls under the forward threshold. In the first scenario, which is assumed to take place if expected supporting evidence is missing, we would expect to find no impact on the fixation times occurring right before a regression is triggered. In the second case, however, which is assumed to be sensitive to integration difficulties, a regression is triggered before the confidence level of word *n+1* reaches the forward

[23] Note that there is evidence by MacGregor and colleagues (MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012) showing that different cortical activations for words vs. pseudowords can be observed very rapidly (after about 50 ms). However, the retrieval of all lexical information associated with a visual string needs time.

threshold. Accordingly, we would expect shorter fixation times if a regression is triggered in response to integration difficulties. Thus, shorter fixation times before regressions are directly predicted by the architecture of the *Information Gathering Framework*.

Whether this assumption of a distinction between regression types can be supported by empirical evidence, however, has to be a subject for future research.

3.2.5 Different rereading patterns

Von der Malsburg and Vasishth (2013; see chapter 2.4.3) investigated the regression pattern when reading locally ambiguous sentences. In particular, they were able to identify three different patterns: checking, rereading and rapid regressions to the pre-verbal region. Furthermore, the authors reported that readers often had a preference for one of these regression patterns. These findings raise two important questions: First, why do different reading patterns exist? Second, why do exactly these three patterns exist?

With regard to the both questions, we must initially remark that the definition of a regression pattern is not fixed. Thus, the categorization of different regressions to patterns and the definition of exactly three patterns is (at least to some extent) arbitrary and was subject to the decision of the authors. In addition, we have to say that the investigations of von der Malsburg and Vasishth do not provide an appropriate testing ground for the *Information Gathering Framework* due to methodological issues (e.g., they only report regressions initiated in the disambiguating and post-disambiguating region; they always used very similar sentence types; the ambiguous region is still very close to the disambiguating region; and so forth). Nonetheless, we think that these results in general fit well with the current framework.

First, within the *Information Gathering Framework* it is assumed that the majority of regression targets are selected on the basis of strategies and these strategies should in turn reveal patterns of regression targets as opposed to highly variable distributions of landing sites. Thus, the findings of patterns supports the assumption that regression targets are selected by a strategy.

Second, the three reported regression patterns also fit well with the assumptions of the proposed framework. The checking pattern could be a plausible response to integration difficulties where the production rules reveal that more evidence about the disambiguating region is needed. The rapid regressions to the pre-verbal region, on the other hand, provide a plausible response to cases where additional evidence is missing (without integration difficulties). Finally, the rereading pattern, as already discussed in chapter 3.1.9, provides a useful strategy if the source of an error cannot be determined, either due to missing linguistic knowledge or due to low memory capacities.

Third, as already discussed above, the finding that readers showed preferences for certain regression patterns fits well with the assumption that the regression target selection is subject to individual differences. These differences occur due to individual language knowledge and exposure and individual memory resources.

In sum, the findings of von der Malsburg and Vasishth are perfectly in line with the *Information Gathering Framework*, although more research is needed to actually test the predictions of the model.

3.2.6 Regressions in child and adult readers

Another interesting issue with regard to regressive eye movements is the development of reading behavior over a lifetime. Since reading is an acquired skill that has to be practiced, differences between readers who just started to read and those who already have many years of experience are to be expected.

Buswell (1922; see also Taylor, Frackenpohl, & Petee, 1960) was the first to report systematic differences between child and adult readers. He found that during the first six years of school, the mean fixation duration and the number of fixations decreases. At the same time, no clear trend for a reduction of regressions was indicated by the results. This pattern has been confirmed by subsequent research (e.g., Rayner, 1985). A more fine grained analysis, however, revealed that intra-word regressions become less frequent, but that the number of inter-word regressions increases from child to adult readers (McConkie et al., 1991). Furthermore, there is evidence that the perceptual span for developing readers is smaller than for adult readers (Rayner, 1986).

These findings can also be easily implemented in the *Information Gathering Framework* because the model provides some room for the impact of language and reading experience.

First, longer and more fixations for developing readers are expected because during a fixation the visual input is received and the lexical quality level is computed. Since word representations are sensitive to frequency, a child needs more time to reach a certain lexical quality level. This in turn affects the computation of the confidence level, so that the forward threshold which triggers the saccade to the next word is reached later.[24] In addition, the computation of the confidence level may take more time because the repertoire of production rules is smaller, thus leading to increased integration difficulties.

Second, the increase of inter-word regressions for adult readers can be explained by setting a lower first-pass threshold which causes lower quality levels to be assessed as sufficient. As a consequence, the probability that additional evidence about a word is required increases, so that in response, a regression has to be performed. This typical accuracy-speed tradeoff phenomenon points to the fact that more speed in general corresponds with less accuracy. In the case of reading strategies, developing readers obviously start with a strategy that aims for high accuracy but with the disadvantage of a very slow reading speed. Adult readers, by contrast, tend to use a more risky strategy, which is faster but also leads to more regressions. This might be because adult readers are better able to compensate for information deficits by using linguistic and world knowledge.

Recall the simulations of Bicknell and Levy (2010), who investigated the interaction between the forward and the backward threshold in their model of falling confidence. The authors provided evidence that from a mathematical point of view, that the most effective reading strategy (the one that is fast and leads to a high accuracy at the same time) is one that allows for a more risky first pass reading with making occasional regressions. Thus, the finding that more risky reading strategies develop over one's lifetime is also well justified by theoretical considerations (see also, for example, Rayner, Yang, Castelhano, & Liversedge, 2011, and Risse & Kliegl, 2011, for corresponding evidence from reading experiments). This fits well with the assumption that predictions (that are based on the context and on world knowledge) may cause that a lower level of confidence is to be assessed as sufficient, which leads to shorter first-pass reading times. Note, however, that within the *Information Gathering Framework* a lower forward threshold does not necessarily lead to more regressions, because the forward and backward threshold are assumed to be independent, although they often highly interact.

---

[24] Note that the *Information Gathering Framework* is based on words and thus inter-word saccades, so that several fixations within a word are viewed as one long fixation. Although this is a simplification, of course, we think that it is a reasonable simplification for the current purpose because intra-word saccades increase the lexical quality of the same word which in turn affects the confidence level. More fixations for developing readers (which are in fact primarily more fixations within a word) can be therefore viewed as increasing lexical quality.

3.3 Predictions of the *Information Gathering Framework*

In the last chapters we have seen how the *Information Gathering Framework* may account for a variety of empirical findings reported in the literature. Another important factor for the strength of a model, however, is that it allows for further predictions. In the following, we will therefore discuss several more predictions that can be derived from the architecture of the model and that have to be tested by future research. But note that not all predictions discussed here will potentially falsify the model. For example, the *Information Gathering Framework* assumes that new input is matched against predictions arising from previous input, which is one of the core principles of the model. If we were to find empirical evidence against this assumption, this would seriously question the validity of the model. But whether these predictions are accomplished on the basis of production rules, by contrast, does primarily affect the detailed architecture of the model but not its core principles.

3.3.1 Regressions targets within and outside of the perceptual span

The *Information Gathering Framework* makes a strong prediction with regard to the target selection of regressions: Only words within the perceptual span, which is assumed to comprise about 15 characters to the left of the current fixation, can be selected by an explicit linguistic computation as a regression target. Words outside of the perceptual span are assumed to only be selected by a backward strategy. This division should be reflected by the empirical data somehow.

First, it would be a quite unexpected finding if the regression landing sites show, for example, a Gaussian or a linear distribution over the sentence, thus ranging from very short to very long sizes with no further distinctions. We would rather expect that the majority of regressive saccades land within the perceptual span[25]. In addition, we would expect that we are able to find a clear pattern for regressions that land outside the perceptual span because these regression targets are assumed to be selected by a strategy. And a regression strategy (even if more than one exists) should not lead to a random

[25] Note that we assume that all words within the perceptual span could be used for the computation of the regression target. This also implies that the actual regressive saccade could be longer than 15 characters because parts of the target word could be still within the 15 character window which allows for the retrieval of the lexical information of the whole word. This, though, makes the hypothesis harder to test.

distribution of landing sites but to landing site patterns, which in turn help us to identify the applied strategies.

Second, in the case that there exists a well-defined target position from a theoretical linguistic point of view (as for example, in garden path sentences), we would expect that this defined target position is selected as a regression target only if it is within the perceptual span. If the ambiguous word is outside the perceptual span, for instance, no preference for a selection of this word is predicted, unless it is selected by the strategy[26].

Note that from the experiments investigating regression target positions that were discussed in chapter 2.4 (Frazier & Rayner, 1982; Mitchell et al., 2008; Meseguer et al., 2002), none could really help us to test this hypothesis. The experiments by Frazier and Rayner as well as by Meseguer and colleagues used sentences where the ambiguous region was very close to the disambiguating region, so that the ambiguous region was still within the perceptual span when the integration difficulties appeared. The study carried out by Mitchell and colleagues, by contrast, was even designed to test potential influences of the distance between the ambiguous and disambiguating region, but due to a lack of statistical power the authors did not find clear results. Thus, it is still an open question if the prediction made by the *Information Gathering Framework* with regard to the perceptual span can be verified by empirical evidence.

3.3.2 Shorter fixation durations before regressions due to integration difficulties but not before regressions due to missing evidence

As we have already discussed in chapter 3.2.4, fixations tend to be shorter when they are followed by a regression compared to cases when they are followed by a progressive saccade. We have also seen how the *information gathering hypothesis* may account for this finding. But our model makes an additional prediction: Fixations should only be shorter when the confidence level of word n falls under the forward threshold during a fixation on word $n+1$, but not if the confidence level of word n is not reaching the backward threshold until the confidence level of word $n+1$ reaches the forward threshold. Thus, we would expect that shorter fixation durations only appear before regressions that are caused by integration difficulties. In the case of regressions that are caused by

[26] It is probably a challenging task to find an experimental setting that allows for a distinction between strategy based regressions and those that are the result of clear linguistic processing.

missing evidence (which for example is assumed for the majority of sentence-final regressions), by contrast, no shorter fixations durations are expected.

3.3.3 Independency of forward and backward threshold

Within the *Information Gathering Framework* it is assumed that the duration of first pass reading times is monitored by the forward threshold and the probability to regress by the backward threshold. Although there is considerable evidence that these two threshold highly interact (as indicated by the speed-accuracy tradeoff), we assume that these two parameters can be set independently.

Thus, we predict that there are cases where a more risky forward strategy does not necessarily lead to an increased probability of regressions. On the other hand, there should be cases where the probability of regressions is increased despite the fact that first pass reading times are not reduced.

3.3.4 Regressions are sensitive to task modulations

Since regressions are assumed to be mediated by the forward and backward threshold, we would expect that an adjustment of these thresholds should have an impact on the probability of triggering a regression. In particular, top-down influences like task or time pressure should affect the regression behavior during reading with leading to more or less regressions respectively.

3.3.5 Regression targets within the perceptual span of children

The perceptual span is assumed to play a crucial part for the selection of a regression target within the *Information Gathering Framework*. Whereas there is only evidence that the size of the perceptual span switches toward the reading direction (Apel et al., 2012), the actual size of this perceptual span for adult readers remains unclear. In the current model we assumed that the perceptual size to the left has a size of 15 characters, according to the size for progressive saccades. Whether this assumption is true, however, has to be tested by future research.

In addition, there is evidence that the perceptual span is smaller for children (see, for example, Sperlich, Schad, & Laubrock, 2015). Thus, we would expect that we may find evidence for this reduced perceptual span also in the selection of target positions for regressions, if the target selection is determined by the perceptual span.

3.3.6 Language specific strategies

One interesting aspect that may play a role in developing backward regression strategies could be the language under consideration. In particular, it has been shown that languages vary in the way they map functions with forms (MacWhinney, Bates, & Kliegl, 1984). In English, for example, the argument that takes the actor role can be most reliably identified on the basis of word order information. In German, by contrast, case marking and animacy play a far greater role in actor identification. Thus, it has been proposed that people rely on certain form cues (like case marking, word order and so on) in order to map arguments to their function in the sentence. Crucially, these cues vary between languages with regard to their strength, and this strength has been established on the basis of experience with this particular language.

If this assumption holds true for sentence interpretation in general, it is also possible that readers establish regression strategies that rely on the properties of the particular language. In German, for example, the important case marking can be often found on the determiner but not on the noun itself. Thus, it would be a useful strategy to increase the quality level of the determiner by a regressive eye movement, because the determiner most likely provides the required information to map the argument to a function. In English by contrast, the determiner contains no case marking information and thus the re-inspection of the determiner should not be very useful in order to solve syntactic integration problems. Accordingly, we would expect that determiners in German are more likely to become the target of a regression than determiners in English.[27]

Unfortunately, to our knowledge no eye tracking corpus neither in English nor in German exists that allows for the examination of this prediction[28]. In addition, because of the differences between languages with regard to sentence structure, lexicon, and so on, it is difficult to create experimental settings that can test this prediction.

[27] It is possible, of course, that the selection of determiners as regression targets in German is the result of a linguistic computation based on the prediction rules and not part of a regression strategy.

[28] Most eye tracking corpora only contain information about single word fixations in first pass reading (first fixation time, skipping rate and so on). The German Potsdam Sentence Corpus (PSC), for example, that consists of 144 German sentences (Kliegl, Grabner, Rolfs, & Engbert, 2004) does not contain any regression data. Thus, it is not possible to assess the landing site distributions of regressive eye movements.

4. Summary

In the last chapters we have reviewed the empirical findings reported in the literature with regard to regressive eye movements during sentence reading and showed how current models of eye movement control (E-Z Reader 10, SWIFT, Model of falling confidence) try to capture these results. It became apparent that none of the models discussed above are able to sufficiently explain what precisely causes a regression during reading. In particular, none of these models provides an answer to the question how a regression target is selected because they all focus on regressions to the immediately preceding word. Thus, they cannot account for the landing site patterns of regressions at all.

The empirical results reported in the literature, however, indicate that the landing site patterns of regressions cannot be explained by a default or random target selection but they also question a strict linguistic guidance that defines a single target position for every regression based on linguistic computations.

Based on these considerations we proposed a new account, the *Information Gathering Framework*, which defines the intake of additional information about previous words' identities as the unifying principle for all regressive eye movements in reading. In particular, the *Information Gathering Framework* assumes that confidence levels for each word are computed and that a regressive eye movement is triggered whenever this confidence level falls under a certain threshold. This could either happen due to integration difficulties or due to missing further evidence. Importantly, the framework tries to model human reading behavior with taking into account the time course of linguistic processing as well as the limited attentional focus. The *Information Gathering Framework* therefore provides some substantial modifications of the FC model proposed by Bicknell and Levy (2010).

Within the *Information Gathering Framework*, the target of a regression is assumed to be selected either on the basis of linguistic computations in the case where the words are within the perceptual span, or alternatively on the basis of experience based backward strategies. Crucially, we have discussed several factors (like linguistic knowledge and memory capacities) that may affect the selection of regression targets and may explain why it is sometimes hard to find a homogenous landing site pattern.

In the last chapters of the previous section, we have seen how the *Information Gathering Framework* may account for several empirical findings that were reported in the literature. In addition, predictions of the model for future research were laid out.

Having provided a theoretical but empirically motivated framework for regressive eye movements in reading, we will now turn to two experiments that were performed in order to test some important predictions of the *Information Gathering Framework* and to shed more light on the role of regressive eye movements in reading. The first experiment focuses on the influence of task manipulations on rereading behavior during sentence interpretation. The second experiment focuses on the neural correlates of regressive eye movements and the impact of experimental settings by using a novel technique, namely combining eye tracking with functional magnetic resonance imaging (fMRI).

Part II: Testing the Information Gathering Framework

In the first part of this thesis we have outlined a new model, the *Information Gathering Framework*, which aims to answer two important research questions:

a) What triggers regressive saccades during sentence reading?
b) What determines the landing position of regressive eye movements?

Although we have shown how the proposed framework may account for several findings reported in the literature, this does not per se prove the validity of the model. In particular, because none of the experiments discussed so far were designed (and analyzed) in order to test the assumptions of the model, nor was the raw data available to us in order to do some post-hoc examinations, the strength of the model has to be further evaluated by testing implicit and explicit predictions that can be derived from the architecture of the model.

The first experiment we will present in the following focuses on the role of the backward threshold for triggering regressions by introducing different reading strategies. This eye tracking experiment also enables us to examine a variety of further predictions of the *Information Gathering Framework*. The second experiment focuses on the neural correlates of regressive eye movements during reading by using concurrent fMRI / eye tracking measures. Since almost nothing is known about regressive eye movements from a neurocognitive perspective, this experiment should be viewed as some pioneering work for future research. We will also provide some ideas how these results can be implemented into the *Information Gathering Framework* which has been basically developed on the basis of behavioral data without making strong assumptions about the neural underpinnings of regressions.

5. Experiment 1

5.1 Introduction[29]

It is often noted that despite its complexity, skilled reading is a fairly automatic process (e.g., LaBerge & Samuels, 1974). Under normal circumstances, the visual, cognitive, and linguistic aspects of processing that are necessary for text comprehension work together seamlessly, with little need for conscious control or strategy. Consistent with this assumption, models of eye movements in reading such as E-Z Reader (Reichle et al., 1998; Reichle et al., 2006) and SWIFT (Engbert et al., 2002; Engbert et al., 2005) have regarded the speed of basic processes of word recognition as the primary determinant of the eyes' forward movement through the text, and have not explicitly provided any role for variability in readers' strategies or goals. And though higher-level processes of syntactic parsing and semantic interpretation also influence eye movements (see, e.g., Rayner & Liversedge, 2011; Staub, 2015, for reviews), these processes are also usually assumed, at least implicitly, to operate without substantial variation based on, for example, the level of understanding that is required of the reader.

At the same time, several studies have examined effects of explicit task manipulations (e.g. proofreading vs. reading for comprehension) on eye movements in reading. These studies have shown that eye movement measures like fixation duration, skipping rate and probability of regressions can be modulated by specific reading tasks (e.g., Kaakinen & Hyönä, 2010; Schotter et al., 2014; Radach, Huestegge, & Reilly, 2008). Arguably, however, these studies may not be informative about how eye movements in reading are modulated by more typical variations in the demands of the reading situation. There are situations in which the reader needs to extract only the 'gist' of the text, but there are also situations in which the reader must be sure to encode information in relatively fine detail. However, there is little prior research investigating how implicit modulations of reading strategy or depth of processing influence eye movements.

To our knowledge, only a single study (Wotschack & Kliegl, 2013; hereafter WK) has addressed the question of whether, and how, the difficulty of comprehension ques-

[29] Large parts of the data we will present here (but not the word-based analysis of the reading data) have been already published in the following article: Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2017.

tions asked after sentences might influence eye movements more generally. WK manipulated comprehension demands between subjects, by means of presenting either easy or difficult comprehension questions after sentences from the Potsdam Sentence Corpus (Kliegl, Grabner, Rolfs, & Engbert, 2004; Kliegl, Nuthmann, & Engbert, 2006). Focusing on word-based eye movement measures, WK found that among younger readers, difficult comprehension questions increased the frequency of regressive eye movements, and as a result had a pronounced effect on total reading time. However, they did not find reliable effects on other measures such as skipping probability and first fixation duration. (For a group of older readers, with a mean age of about 70 years, the question manipulation did influence these measures.)

Crucially, the finding that comprehension demands may modulate reading behavior and especially regressions has important implications for the *Information Gathering Framework* outlined in the first part of this thesis. Whereas the impact of reading strategy and required processing depth on eye movements during reading is not taken into account by models like E-Z Reader and SWIFT, the architecture of the *Information Gathering Framework* is based on the assumption that the amount of information that is assessed to be sufficient for the current purpose varies among readers and especially among tasks. This amount of information in turn determines the execution of progressive but also of regressive saccades during reading. Thus, the *Information Gathering Framework* explicitly predicts that top-down influences like task or even comprehension questions should affect eye movements during reading.

In particular, there are two mechanisms within the model that are assumed to be sensitive for the current reading goal: The forward threshold that determines the level of confidence into a word's identity that is assessed to be sufficient for the current purpose and in response initiates a saccade to the next word, and the backward threshold that regulates the level of confidence that causes a regressive eye movement. Both thresholds are assumed to be highly interactive but can be set individually. If this assumption holds true and if we were able to adjust only the backward threshold but not the forward threshold by a certain task, the architecture of the *Information Gathering Framework* would lead us to two very important predictions:

First, it should be possible to increase the probability of regressive eye movements without increasing or decreasing first pass duration because these two parameters are independent (see chapter 3.1.6). Thus, if the backward threshold is set higher which would lead to more regressive eye movements, this does not automatically affect first pass reading time since the execution of progressive eye movements is monitored by the forward threshold.

Second, adjusting the backward threshold should increase the number of regressions due to missing evidence but not due to integration difficulties. For this prediction recall the two different scenarios for triggering a regression. In the first scenario, the confidence level of word *n* falls under the forward threshold (during a fixation on word *n+1*) because post-lexical processing reveals integration difficulties. In this case, a regression is triggered independent of the backward threshold. In the second scenario, the confidence level of word *n* does not reach the backward threshold before the confidence level of word *n+1* reaches the forward threshold which is assumed to happen if expected evidence is missing. In this case, the probability to reach the backward threshold within the given time interval (i.e., until the confidence level of word *n+*1 reaches the forward threshold) depends on the level of the backward threshold: The higher the backward threshold, the higher the probability that a regression is triggered because the confidence level is not reaching the threshold. Thus, the *Information Gathering Framework* predicts that an adjustment of the backward threshold should only affect the probability of making regressions due to missing evidence but not due to integration difficulties.

In order to test these predictions, we conducted an experiment that followed up on WK's results in several ways. Since comprehension questions aim to probe higher-order language processing, we assessed them to be an appropriate tool to adjust the backward threshold. In contrast to the forward threshold which is very likely to be affected by tasks that focus on low-level linguistic processing (like proof-reading), the backward threshold should be sensitive to tasks that require syntax and semantic integration. In addition, different levels of processing depth which are elicited implicitly by different comprehension questions are more likely to reflect natural reading situations than explicit tasks do. Thus, a manipulation of the difficulty of comprehension questions seems well suited to test the predictions of the *Information Gathering Framework* which aims to model real-world reading processing instead of artificial reading tasks.

In order to test the predictions of the *Information Gathering Framework* we nonetheless did a few more substantial modifications on the study design used by WK. The first question we addressed is whether more pronounced effects of question difficulty may appear in the eye movement record with a more pronounced difficulty manipulation. The limited influence of the question manipulation on young adults' reading behavior in WK's study may be due to the fact that the 'difficult' comprehension questions in that study were not, in fact, very difficult. The young adult readers in the difficult question condition tested by WK achieved 95% accuracy on the comprehension questions, compared to 97% in the easy condition; the high level of accuracy in the difficult condition is especially striking given that WK used three-alternative multiple choice questions.

Second, our experiment allows a more fine-grained analysis of the linguistic processes that trigger regressions by using sentence material with well-controlled syntactic structures, including anomalous sentences. This enables us to assess the location in the sentence from which regressions are initiated when readers are confronted with difficult comprehension questions, but also makes it possible to distinguish regressions that are triggered in response to integration difficulties from regressions triggered by missing further evidence. This is an important prerequisite in seeking to test the predictions discussed above.

Finally, the present experiment also aims to add some more insights to an important question of language processing, namely, whether a global manipulation of question difficulty would specifically impact how readers deal with anomalous content. In recent years, the notion that language processing is merely 'good enough' for the task at hand has received extensive discussion (e.g., Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; see also Christianson, 2016, which introduces a special issue of the *Quarterly Journal of Experimental Psychology* on the topic of 'good enough' processing). It has been suggested that comprehenders may often impose an interpretation that is consistent with real-world knowledge rather than with the literal meaning of the text, when such an interpretation suffices for the task at hand and when the real meaning conflicts with heuristic principles of sentence comprehension, e.g. when a sentence does not adhere to the expected Agent-Action-Object template in English. In the present experiment, we ask whether readers who are presented with only easy comprehension questions would be more likely to overlook a reversal of thematic roles (e.g. *the flower is drawing the girl*), and whether they would be especially likely to do so when the subject and verb are highly lexically associated (e.g.*, the flower is picking the girl*). Specifically, we test the hypothesis arising from the good enough approach that the anomaly effect on first pass reading of the verb and object in these sentences may be reduced when the reader receives only easy questions, and that this effect may be reduced especially when there is a high degree of association between the subject and verb. Processing of these so-called 'semantic reversal anomalies' (hereafter SRAs) has been investigated using Event-Related Potentials (ERPs; e.g. (Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003, among many others), but has not been previously investigated using eye movements in reading.

5.2 Method

5.2.1 Participants

Ninety-two undergraduate students at the University of Massachusetts, Amherst participated in the experiment (45 in the easy condition). Two additional participants were excluded due to poor accuracy on comprehension questions in the easy condition (<80%, which is > 2.5 SD from group mean) and three were excluded due to very slow average total sentence reading time (>2.5 SD from group mean). All participants had normal or corrected-to-normal vision and were native speakers of American English. All were naive concerning the purpose of the experiment and received course credit for participation.

5.2.2 Materials

The materials in the experiment as a whole consisted of three sets of sentences, which we refer to as semantic reversal anomalies (SRA; 36 sentences); relative clause sentences (RC; 39 sentences), divided between sentences that contained object and subject relative clauses, as described below; and garden path sentences (GP; 24 sentences).[30] However, only the SRA sentences were manipulated within subjects and within items; all subjects read the same versions of the RC and GP sentences. All RC and GP sentences were followed by comprehension questions, but because half of the SRA sentences were anomalous, and because different questions would be necessary for the different versions of each item, we did not include comprehension questions for these items. In total, questions followed 63 of the 99 sentences in the experiment. The SRA sentences were allocated to 4 experimental lists according to a Latin Square design and randomly intermixed with the RC and GP items. Question difficulty was manipulated between subjects, and participants were randomly assigned to one of the 4 lists and one of the two question difficulty conditions. Thirty-three of the difficult questions and 32 of the easy questions required a YES answer.

[30] The number of SRA items was motivated by the 2 x 2 factorial design of that subexperiment, with the 36 items allowing each subject to read nine sentences in each condition. However, the number of RC and GP sentences was determined simply by the ease of constructing these sentences and the general desire to avoid too many repetitions of the same structure; for these sentence types, there was no need to, e.g., use an even number of items.

An example SRA stimulus set is given in Table 1. We manipulated the factors ANOMALY (non-anomalous vs. anomalous) and ASSOCIATION (high vs. low), in a 2 x 2 design. Non-anomalous sentences had an animate subject and an inanimate object. In anomalous sentences, the subject and object were reversed. The verbs required an animate subject, so that the reversed sentences were anomalous at the point of reaching the verb. The verb used in the high association condition (e.g., *picking*) was judged to be strongly associated with the inanimate object (e.g., *flower*) based on a pre-test, described below. Note that verbs were used in the present progressive tense, as opposed to the past tense, to avoid the possibility that subjects would initially adopt a reduced relative clause reading at the verb (e.g., *The flower picked by the girl was...*).

Table 1: *Sentence types used as stimuli in the experiment. Sentence regions for the region-based eye-tracking analysis are marked by the vertical strokes.*

1. Semantic Reversal Anomalies (SRA)			
a)	On a sunny afternoon | the girl | is picking | the flower | for the dining table.		*non anomalous, high associated*
b)	On a sunny afternoon the girl is drawing the flower on a little sketchpad.		*non anomalous, low associated*
c)	On a sunny afternoon the flower is picking the girl for the dining table.		*anomalous, high associated*
d)	On a sunny afternoon the flower is drawing the girl on a little sketchpad.		*anomalous, low associated*
2. Relative Clause Sentences (RC)			
a)	The chef | that distracted the waiter | sifted the flour onto the counter.		*subject relative (SRC)*
	I)	Did a chef do something?	*easy*
	II)	Did the waiter distract the chef?	*difficult*
b)	The executives | that the lawyers sued | roused themselves from slumber.		*object relative (ORC)*
	I)	Did a policeman do something?	*easy*
	II)	Was it the executives who roused themselves?	*difficult*
3. Garden Path Sentences (GP)			
	John borrowed | the rake or the shovel | turned out to be sufficient.		
	I)	Is there a shovel?	*easy*
	II)	Might the rake have been borrowed?	*difficult*

The high and low associated verbs in the sentences did not differ significantly in mean length in characters (high 7.4, low 7.5; $t(35) = -0.36$, $p=0.72$) or mean frequency (log Subltex: Brysbaert & New, 2009; frequency 2.6 for high, 2.5 for low; $t(35) = 0.27$, *p*=0.79). The inanimate nouns were slightly shorter on average than animate nouns (animate 6.6, inanimate 5.5; $t(35) = 2.48$, $p<0.05$) and were also slightly more frequent (animate 2.8, inanimate 3.2; $t(35) = -2.25$, $p<0.05$).

In an online pre-test on Amazon Mechanical Turk, verb and noun pairs were assigned to 4 lists according to a Latin Square design. For each pair of words the participants were asked to give a rating on a 5-point scale of how strongly associated the meanings of the words are (1 – low associated, 5 – high associated). Each list contained 55 word pairs in random order and was rated by 13 participants who self-reported as native speakers of American English. The questionnaire took between 5 and 10 minutes and participants received $1. Twenty-four additional word pairs were rated by 8 further subjects in a paper-and-pencil questionnaire.

Based on the results from the pre-test we selected 36 stimuli for which the association ratings between the verb and the two inanimate nouns were very different. The difference between them was highly significant for both animate (high associated 3.95, low associated 2.25; $t(35) = 7.53$, $p<.001$) and inanimate nouns (high associated 4.49, low associated 1.92; $t(35) = 24.99$, $p<.001$).

In a second questionnaire on Mechanical Turk we asked participants to rate the naturalness of the entire sentence (1 – very unnatural, 5 – very natural). The critical sentences were presented as they would appear in the main experiment and were again allocated to 4 lists. Each list contained 55 sentences in random order and was rated by 13 further participants. Each questionnaire took about 10 minutes. Participants received the same remuneration for their participation. Again, 24 additional sentences were rated by 8 further subjects in a paper questionnaire. The ratings of the 36 experimental stimuli that were selected based on the association criteria described above suggest no influence of ASSOCIATION either within the anomalous (high associated 1.25, low associated 1.29) or the non-anomalous sentences (high associated 4.27, low associated 4.12) but a clear effect of ANOMALY. This pattern was confirmed by a repeated measures by-items ANOVA with ANOMALY and ASSOCIATION as fixed effects that showed a significant main effect of ANOMALY ($F_2(1,35) = 758.20$, p<.001), but neither a main effect of ASSOCIATION ($p=.59$) nor an interaction ($p=.11$).

The 39 RC sentences were followed by either easy or difficult comprehension questions, depending on the task condition. Two examples are shown in Table 1. Seventeen of these sentences contained a subject-modifying subject relative clause (SRC; 2a), and 22 contained a subject-modifying object relative clause (ORC; 2b). Object relative clauses in particular are known to induce measurable processing difficulty in the eye movement record (e.g., Staub, 2010). In all cases, the difficult question required the reader to correctly assign thematic roles to the noun phrases in the sentence, while the easy question did not.

Finally, an example of the 24 GP sentences is also shown in Table 1, with the corresponding easy and difficult comprehension questions. The GP sentences always involved a temporary ambiguity between a noun phrase coordination structure and a clausal coordination structure, which was ultimately resolved toward the (initially dispreferred) clausal coordination analysis (e.g., Staub & Clifton, 2006; Staub, 2007). A comma after the initial clause (e.g., after *rake* in the example in Table 1) would avert the garden path, but the comma was never present. The difficult question always required the reader to successfully interpret the thematic roles of the noun phrases in the sentence, while the easy question did not.

5.2.3 Procedure

Eye movements were recorded using an SR Research EyeLink 1000 tracker with a sampling rate of 1000Hz. The viewing was binocular, but only one eye's movement was monitored. In most cases this was the right eye but due to technical problems in some cases the left eye was tracked. All sentences were displayed on one line on a CRT monitor 55 cm from the participant, in 12 point Monaco font. At this distance, three characters corresponded to approximately one degree of visual angle; the resolution of the eyetracker was less than one character.

Participants were asked to read for comprehension and were told that after some of the sentences they would be required to respond to a comprehension question presented on the screen by pressing one of two buttons on a gamepad. They were also told that some of the items might be 'a little weird'. Once the participant was seated at the eyetracker, the tracker was aligned and calibrated in a single line calibration. The experiment began with six practice trials, and took about 25 minutes in total.

5.2.4 Analysis

The main goal of the current experiment was to investigate the impact of comprehension difficulty on eye movements when reading anomalous and non-anomalous sentences.

But in addition, we were interested in the general pattern of all regressive eye movements that were performed during the experiment, especially when reading the SRA sentence. Thus, we carried out two different analyses.

The first analysis focused on sentence reading and computed eye movement measures for each sentence region separately (region-based analysis). For this analysis, we deleted all trials with first-pass blinks or track losses in the verb region because the verb region of the SRA sentences was of critical interest as the location in which an implausibility arises. This resulted in excluding about 11% of these sentences. No RC or GP sentences were excluded due to blink or track loss, as there was no pre-identified critical region of these sentences. The initial regions for analysis were defined as given in Table 1; below we discuss results with an alternate regioning scheme.

Eye movement analysis focused on first-pass time (the sum of all fixation durations on a region before leaving it to the left or right) and go-past time (the sum of all fixation durations from first entering a region until leaving it to the right, including any regressive rereading), as well as the probability of a first-pass regressive eye movement from a region. These three measures have the ability to clearly capture potential effects of the task difficulty manipulation on incremental processing of the sentence, and together to distinguish first-pass reading effects from rereading effects. We calculated linear mixed-effects models (for first pass time and go-past time) and logistic regression models (for probability of regressions out) for each region, for each sentence type, using the lme4 package (Version 1.1-5; Bates, Mächler, Bolker, & Walker, 2014) and the NLopt nonlinear-optimization package (http://ab-initio.mit.edu/nlopt) for R (Version 2.15.1, The R Foundation for Statistical Computing, 2012). Following Barr et al. (Barr, Levy, Scheepers, & Tily, 2013), we used the maximal random effect structure justified by our design. These models included random intercepts for subjects and items, and random by-item slopes for the question difficulty manipulation. For the SRA models, random by-subject and by-item slopes were also included for the effects of anomaly and association, and their interaction. Following convention, we treat $|t|$ or $|z| > 2$ as significant.

In order to examine the pattern of regressive eye movements in more detail we carried out a second analysis to which we refer to as the word-based analysis. Because only the SRA sentences allowed for a comparison of regressions in the context of anomalous as well as non-anomalous sentence structures, we restricted the analysis to these sentences only. In contrast to the region-based analysis we identified every single saccade, irrespective of whether it occurred in first pass or second pass reading or whether it went out of a particular region or not. But we excluded saccades that were triggered right before or after blinks as well as the last saccade in a sentence. Since we are interested in inter-word saccades (due to the word-based approach of the *Information Gathering Framework*), we in addition excluded all saccades that landed within the same word (intra-saccades). In sum, 31671 progressive and 10129 regressive saccades entered the word-based analysis.

5.2.5 Results

5.2.5.1 Question Accuracy

On average, question accuracy in the easy condition was 93.57% (after exclusion of two subjects, as noted above; range 80.30% – 100%), whereas for the difficult group average accuracy dropped to 82.70% (range 63.63% – 98.48%); see Figure 6. The difference between conditions was significant ($t(69.99) = -7.71$, $p<0.001$) and was apparent for both kinds of RC sentences (SRC: easy 93.99%, difficult 82.48%, $t(84.68) = -5.22$, $p<0.001$; ORC: easy 92.53%, difficult 77.72%, $t(60.21) = -6.59$, $p<0.001$) as well as the GP sentences (easy 96.11%, difficult 87.94%, $t(95.37) = -4.78$, $p<0.001$).

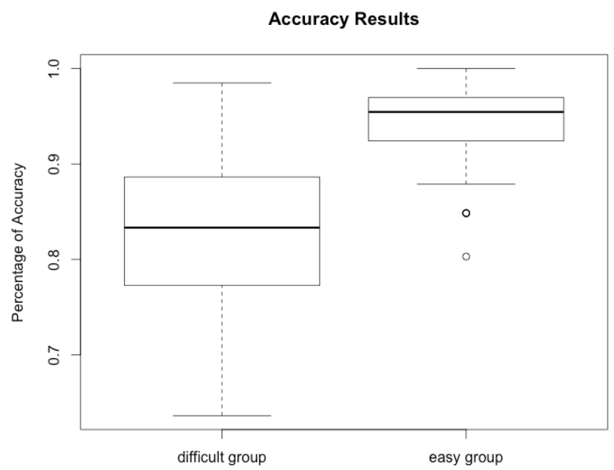

Figure 6: *Comprehension accuracy, by question difficulty condition (re-print with permission from Francis & Taylor).*

5.2.5.2 Analysis 1: Region-based eye movement measures

Semantic Reversal Anomalies

Tables 2 and 3 provide mean reading times for the SRA sentences and statistical results, respectively, for each of the sentence regions. In line with previous research indicating that anomaly has immediate effects in the eye movement record (Rayner, Warren, Juhasz, & Liversedge, 2004), anomaly increased first pass time and go-past time on the verb and object regions (regions 3 and 4), and the probability of a regression from each of these regions. The effect of anomaly also reached significance in go-past time and regressions out for the sentence-final region 5. The main effect of association was significant only in first pass time on the object and in go-past time on the final region, but there was a significant anomaly x association interaction in first pass and go-past time on region 3 and 4. There was a reduced anomaly effect in the associated conditions in region 3, but an increased effect of anomaly with association in region 4. We discuss this pattern in more detail in the Discussion.

The question difficulty manipulation had a significant effect on go-past time and regressions out for the sentence-final region, with longer reading times and more regressions with difficult questions. However, there were no significant effects of this manipulation on the earlier regions. In particular, question difficulty did not modulate the size of the anomaly effect on the critical regions 3 and 4, as would for example be indicated by reduced first pass reading times.

Table 2: *By-subject means and standard deviations (in parentheses) of first pass time, go-past time and probability of regressions out for semantic reversal anomalies (SRA). Abbreviations: ano = anomalous sentences, non ano = non-anomalous sentences, easy = easy comprehension questions, difficult = difficult comprehension questions, low = low associated, high = high associated.*

	Region 1 *On the sunny afternoon*				**Region 2** *the flower / the girl*				**Region 3** *is picking / is drawing*				**Region 4** *the girl / the flower*				**Region 5** *for the dining table*			
	easy		difficult		easy		difficult		easy		difficult		easy		difficult		easy		difficult	
	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high
First pass time																				
ANO	526 (190)	514 (151)	525 (128)	520 (173)	380 (114)	379 (105)	390 (110)	356 (90)	430 (122)	400 (90)	422 (112)	398 (101)	402 (95)	402 (100)	395 (94)	401 (94)	897 (269)	836 (205)	860 (274)	856 (307)
NON ANO	507 (157)	497 (128)	530 (183)	547 (172)	382 (122)	393 (105)	384 (114)	385 (88)	372 (97)	371 (82)	360 (81)	368 (101)	354 (87)	324 (73)	373 (79)	345 (77)	903 (288)	892 (247)	893 (272)	881 (238)
Go-Past time																				
ANO	549 (201)	518 (152)	534 (127)	528 (171)	457 (168)	443 (125)	475 (147)	453 (155)	516 (160)	492 (145)	526 (131)	499 (158)	556 (174)	597 (229)	593 (172)	599 (185)	1830 (620)	1655 (647)	2269 (737)	2323 (1144)
NON ANO	519 (160)	503 (137)	540 (189)	555 (167)	494 (147)	487 (165)	483 (170)	486 (146)	437 (114)	426 (118)	424 (146)	444 (133)	419 (138)	359 (94)	430 (113)	409 (162)	1491 (486)	1377 (497)	1886 (731)	1753 (605)
Regressions Out																				
ANO					11.11 (12.70)	12.62 (13.14)	13.27 (15.61)	15.32 (17.88)	12.97 (11.49)	12.72 (11.21)	16.65 (14.77)	16.70 (16.43)	22.31 (16.60)	25.47 (17.93)	23.54 (16.18)	27.24 (18.49)	58.21 (27.31)	54.88 (26.77)	74.56 (21.81)	64.81 (28.05)
NON ANO					17.16 (19.18)	12.92 (14.10)	14.97 (15.23)	16.76 (15.61)	10.41 (11.66)	8.34 (9.73)	9.15 (10.28)	10.19 (10.17)	10.34 (11.50)	6.67 (10.02)	9.85 (10.72)	10.00 (11.89)	49.98 (30.59)	47.43 (31.44)	60.69 (26.30)	63.10 (27.25)

Table 3: *Results of the mixed models analysis (first pass time and go-past time) and the logistic regression models (regressions out) for semantic reversal anomalies (SRA).*

Semantic Reversal Anomalies	**Region 1**			**Region 2**			**Region 3**			**Region 4**			**Region 5**		
	On the sunny afternoon			*the flower / the girl*			*is picking / is drawing*			*the girl / the flower*			*for the dining table*		
	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value
First pass time															
Intercept	6.06	0.07	88.74	5.82	0.03	205.84	5.86	0.02	239.38	5.82	0.02	272.46	6.59	0.04	160.82
Anomaly	0.01	0.02	0.31	-0.02	0.04	0.64	0.10	0.02	**4.63**	0.11	0.03	**4.37**	-0.04	0.03	1.23
Association	0.01	0.02	0.60	0.00	0.02	0.08	-0.02	0.02	0.86	-0.04	0.02	**2.28**	-0.02	0.05	0.44
Difficulty	0.05	0.05	0.89	0.00	0.04	0.09	-0.02	0.04	0.43	0.02	0.04	0.46	-0.02	0.05	0.31
Anomaly x Association	-0.04	0.04	0.89	-0.03	0.04	0.75	-0.09	0.03	**2.78**	0.08	0.03	**2.36**	-0.02	0.05	0.37
Anomaly x Difficulty	-0.06	0.04	1.29	-0.01	0.04	0.36	-0.02	0.03	0.45	-0.04	0.04	1.24	-0.04	0.05	0.83
Association x Difficulty	0.02	0.05	0.48	-0.05	0.04	1.39	0.03	0.04	0.70	0.00	0.03	0.08	0.05	0.05	0.92
Anomaly x Association x Diff	-0.02	0.09	0.23	-0.07	0.08	0.95	-0.02	0.07	0.35	0.01	0.06	0.23	0.11	0.10	1.08
Go-Past time															
Intercept	6.11	0.07	90.13	6.01	0.03	182.89	6.01	0.03	201.66	6.03	0.03	224.30	7.31	0.04	171.87
Anomaly	0.00	0.02	0.17	-0.06	0.04	1.31	0.14	0.03	**4.45**	0.32	0.04	**9.03**	0.18	0.03	**5.95**
Association	-0.01	0.02	0.53	-0.01	0.02	0.28	-0.02	0.02	1.14	-0.04	0.03	1.31	-0.07	0.03	**2.28**
Difficulty	0.03	0.05	0.64	0.00	0.05	0.04	0.00	0.05	0.03	0.03	0.05	0.71	0.21	0.07	**2.96**
Anomaly x Association	-0.02	0.04	0.64	-0.01	0.04	0.39	-0.07	0.04	**2.04**	0.12	0.05	**2.71**	-0.04	0.05	0.80
Anomaly x Difficulty	-0.05	0.03	1.47	0.03	0.03	0.82	0.02	0.03	0.54	-0.03	0.04	0.64	0.02	0.06	0.31
Association x Difficulty	0.06	0.04	1.30	-0.02	0.03	0.68	0.04	0.04	0.94	0.01	0.04	0.16	0.05	0.04	1.34
Anomaly x Association x Diff	-0.02	0.07	0.31	-0.09	0.07	1.30	-0.06	0.07	0.91	-0.08	0.08	1.02	-0.03	0.08	0.37
Regressions Out															
Intercept				-2.13	0.13	16.57	-2.34	0.12	19.01	-2.00	0.09	21.79	0.52	0.14	3.65
Anomaly				-0.06	0.12	0.56	0.59	0.17	**3.39**	1.54	0.14	**10.90**	0.44	0.14	**3.05**
Association				0.00	0.12	0.02	-0.17	0.13	1.30	0.02	0.17	0.10	-0.17	0.12	1.47
Difficulty				0.23	0.21	1.08	0.12	0.18	0.68	0.26	0.17	1.51	0.73	0.28	**2.62**
Anomaly x Association				0.28	0.24	1.18	0.19	0.29	0.67	0.34	0.34	0.98	-0.39	0.20	1.91
Anomaly x Difficulty				0.11	0.26	0.43	0.28	0.27	1.02	-0.31	0.26	1.17	-0.01	0.25	0.03
Association x Difficulty				0.20	0.23	0.89	0.28	0.30	0.94	0.34	0.33	1.04	-0.04	0.21	0.20
Anomaly x Association x Diff				-0.77	0.47	1.63	-0.30	0.53	0.57	-0.66	0.56	1.18	-0.78	0.41	1.91

A follow-up analysis addressed the possibility that participants adopted a fairly superficial reading strategy in the SRA sentences, regardless of task condition, because these sentences were never followed by comprehension questions. This hypothesis predicts that the anomaly effect should decline over the course of the experiment, as the lack of comprehension questions becomes apparent. We computed additional statistical models of first pass and go-past time on the critical regions 3 and 4 that included (centered and scaled) trial order and its interactions with the other variables. Random slopes for trial order and its interactions were also included. No main effects of trial order approached significance, and the only significant interaction was between trial order and anomaly in go-past time on region 4 (β = -.05, SE = .02, $t = 2.48$); the corresponding interactions for first pass time, and the go-past interaction for region 3, all had $|t| < .8$. To test whether the significant interaction in go-past time on region 4 resulted in elimination of the anomaly effect late in the experiment, we computed a model identical to our original model, but restricted it to the second half of the experiment. The anomaly effect was still highly significant (β = .28, SE = .04, $t = 6.78$). Thus, the effect of anomaly was present even late in the experiment on both regions 3 and 4, despite the lack of comprehension questions with the SRA sentences.

Relative Clause Sentences

Mean reading times and statistical results for the RC sentences are given in Tables 4 and 5. Though the analysis does not distinguish between the relative clause types, as these sentences were not matched on factors such as lexical frequency and length, we present descriptive statistics separately to illustrate that the effect of task condition is in fact similar for both types. The results for the RC sentences show the same pattern as the SRA sentences. Question difficulty influenced rereading, indicated by significant effects in go-past time and regressions out in the last region. But again, there were no first pass effects for question difficulty in any region.

Table 4: *By-subject means and standard deviations (in parentheses) of first pass time, go-past time and probability of regressions out for relative clause sentences (RC) and garden path sentences (GP). Abbreviations: easy = easy comprehension questions, difficult = difficult comprehension questions, low = low associated, high = high associated.*

	Region 1		**Region 2**		**Region 3**	
	easy	difficult	easy	difficult	easy	difficult
			First pass time			
garden path	728 (165)	776 (195)	927 (216)	923 (206)	877 (216)	848 (219)
subject relative	413 (90)	423 (94)	885 (244)	854 (170)	1326 (283)	1304 (295)
object relative	415 (98)	441 (106)	915 (261)	905 (189)	1375 (281)	1390 (353)
			Go-Past time			
garden path	747 (172)	793 (205)	1099 (286)	1128 (274)	1634 (451)	2049 (694)
subject relative	420 (89)	434 (93)	1027 (270)	1045 (323)	2188 (618)	2808 (904)
object relative	424 (100)	448 (109)	1072 (299)	1123 (295)	2440 (713)	3071 (1054)
			Regressions Out			
garden path			13.98 (9.66)	16.43 (10.90)	54.17 (23.60)	70.62 (21.90)
subject relative			11.63 (10.90)	13.64 (10.05)	50.20 (27.95)	71.21 (24.66)
object relative			12.83 (10.58)	17.43 (14.49)	55.25 (25.55)	70.29 (23.13)

Table 5: Results of the mixed models analysis (first pass time and go-past time) and the logistic regression models (regressions out) for relative clause sentences (RC) and garden path sentences (GP).

	Region 1			Region 2			Region 3		
	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value
					First pass time				
Garden Path Sentences									
Intercept	6.47	0.05	129.21	6.67	0.05	138.92	6.56	0.06	106.33
Difficulty	0.07	0.05	1.37	0.00	0.06	0.01	-0.03	0.05	0.57
Relative Clause Sentences									
Intercept	5.90	0.04	132.14	6.64	0.03	203.06	7.00	0.03	227.69
Difficulty	0.04	0.05	0.91	-0.01	0.05	0.28	-0.01	0.05	0.27
					Go-Past time				
Garden Path Sentences									
Intercept	6.53	0.05	125.64	6.93	0.04	177.57	7.33	0.05	151.22
Difficulty	0.06	0.05	1.26	0.03	0.05	0.52	0.20	0.07	**2.96**
Relative Clause Sentences									
Intercept	5.95	0.05	131.75	6.88	0.04	195.44	7.75	0.04	205.20
Difficulty	0.04	0.05	0.84	0.02	0.05	0.38	0.21	0.06	**3.30**
					Regressions Out				
Garden Path Sentences									
Intercept				-1.99	0.16	12.30	0.66	0.15	4.33
Difficulty				0.23	0.19	1.24	0.92	0.27	**3.36**
Relative Clause Sentences									
Intercept				-2.04	0.11	19.19	0.69	0.15	4.53
Difficulty				0.28	0.19	1.46	0.98	0.29	**3.38**

Garden Path Sentences

Tables 4 and 5 show mean reading times and statistical results for the GP sentences. As for the SRA and RC sentences, there was no effect of the question difficulty manipulation on first-pass reading, but this factor did influence the probability of making regressions out of the sentence-final region, as indicated by a significant effect on go-past time as well as regressions out.

Alternate regioning of sentences

In the analyses above, the task manipulation reliably influenced reading behavior only once readers reached the final region of sentences. We performed an additional analysis focusing on go-past time (the measure in which task effects most clearly emerged) to further clarify this pattern[31]. One possibility is that task effects on material prior to the final region are weak, and that these effects are not clearly seen when this material is divided into multiple regions. To test this possibility, we combined the original pre-final regions (regions 1–4 for the SRA sentences, and 1–2 for the RC and GP sentences) into a single region. In addition, to test whether task effects specifically emerge only very near the end of sentences, we divided the original final region into two regions.

This regioning included all but the last two words of the sentence in one region, and the last two words in another; if this new final region was less than ten characters in length, we included three words. In sum, this alternate regioning divided the sentences into three regions: a long initial region that consisted of the combined pre-final regions from the original analyses, a short second region that consisted of the first part of the final region in the original analyses, and a new final region that consisted of only the last 2–3 words of the sentence (see Table 6).

Table 6: Examples of alternate regioning of the different sentence types

	initial region	prefinal region	final region
Semantic Reversal Anomalies	*On a sunny afternoon the flower is picking the girl*	*for the*	*dining table.*
Relative Clause sentences	*The chef that distracted the waiter*	*sifted the flour onto*	*the counter.*
Garden path sentences	*John borrowed the rake or the shovel*	*turned out to*	*be sufficient.*

Figure 7 shows go-past means and standard errors by region and sentence type. For the initial combined region in the SRA sentences, there was again a large anomaly effect on go-past time (β = .12, SE = .02, t = 6.61), but no other significant effects or interactions. The effect of the task manipulation did not approach significance (β = .02, SE = .05, t = 0.54). For the RC and GP sentences, there was also no task effect on go-past time on the first region (RC: β = .03, SE = .05, t = .58; GP: β = .05, SE = .05, t = 1.09). For the new second region, in SRA sentences there was again an anomaly effect on go-past time (β = .12, SE = .02, t = 5.20), and again no task effect (β = .03, SE = .05, t = .70), and no other significant effects. There was also no task effect in this region for the other sentence types (RC: β = -.01, SE = .05, t = .24; GP: β = -.04, SE = .05, t = .69). In the new, two- or three-word final region of the SRA sentences, however, there were significant effects of both anomaly (β = .18, SE = .04, t = 4.75) and task (β = .25, SE = .09, t = 2.96),

[31] Since the probabilities for regressions out of the first region cannot by definition be calculated, we think that go-past time is the more appropriate measure to detect rereading processes induced by comprehension question difficulty.

as well as a significant main effect of association (β = -.10, SE = .03, t = 3.27). There were no significant interaction effects. There was also a significant effect of task in the new final region of the other sentence types (RC: β = .37, SE = .09, t = 4.04; GP: β = .28, SE = .08, t = 3.45). In sum, effects of task did not reliably appear until readers reached the last 2–3 words of sentences of all three types. Note that the absolute size of the task effect on go-past time from this final region was similar for all three sentence types, with parameter estimates ranging from .09 to .37 in log units.

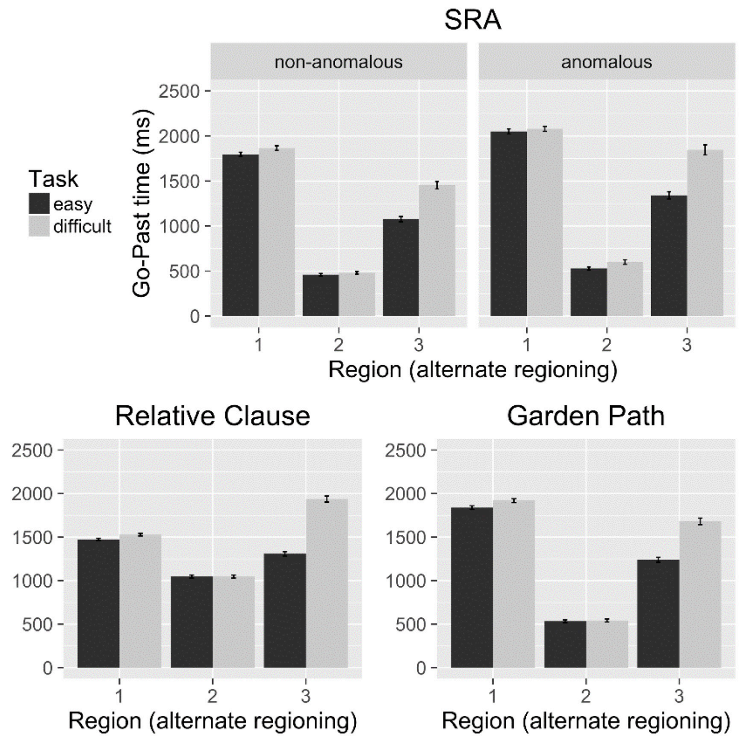

Figure 7: Go-past time means and standard errors by region, based on alternate regioning described in the text, for each sentence type, by task condition (and for SRA sentences, by anomaly condition).

Relation between rereading and question accuracy

Another *post hoc* analysis assessed the relationship between regressions and go-past time on the final region of RC and GP sentences, in the difficult condition, and accuracy on the comprehension questions (see, e.g., Schotter, Tran, & Rayner, 2014, for similar analyses). The previous analyses determined that the difficult task condition influenced reading behavior primarily by increasing rereading from the ends of sentences; this analysis addressed whether such rereading did actually improve comprehension (We did not attempt a similar analysis for the easy condition, as comprehension was near ceiling for the majority of subjects; see Figure 7.). We computed logistic regression models of accuracy that included either regression from the sentence-final region (2–3 final words) or go-past time on this region as a fixed effect, as well as random intercepts for subjects and items and random slopes for the fixed effect. Neither the effect of regression ($\beta = -.23$, $SE = .17$, $z = 1.35$) nor the effect of go-past time ($\beta = -.000066$, $SE = .000050$, $z = 1.33$) reached significance.

5.2.5.3 Analysis 2: Word-based analysis of regressive eye movements

The word-based analysis of regressive eye movements focused on two questions:

a) Does the sentence structure (anomalous vs. non-anomalous) or the task manipulation affect the landing position of regressive eye movements?
b) Does the saccade type (progressive vs. regressive) affect the duration of the fixation before the saccade is triggered?

Distribution of regression landing sites

In the analysis so far, we have not addressed the question of the position at which regressions in the sentence land, as well as how the landing positions differ between sentence types that induced higher-order linguistic difficulties (i.e., anomalous vs. non-anomalous sentences), and between tasks (i.e., easy vs. difficult comprehension questions). In order to investigate the landing site distributions in more detail, we first calculated the length of each regressive saccade in characters.

If we visually inspect the distribution of saccade length for regressions, we see that the majority of saccades are not more than 15 characters in length (see Figure 8). A further investigation revealed that 51.61% of all regressions that were initiated in the final region had a length of 15 characters or less (with a peak for saccades of 6 characters length). For all regressions this proportion even increased to 74.81% which is probably due to the reduced distance from the sentence beginning. However, the number of saccades with a length of more than 15 characters does not further decrease and remains stable.

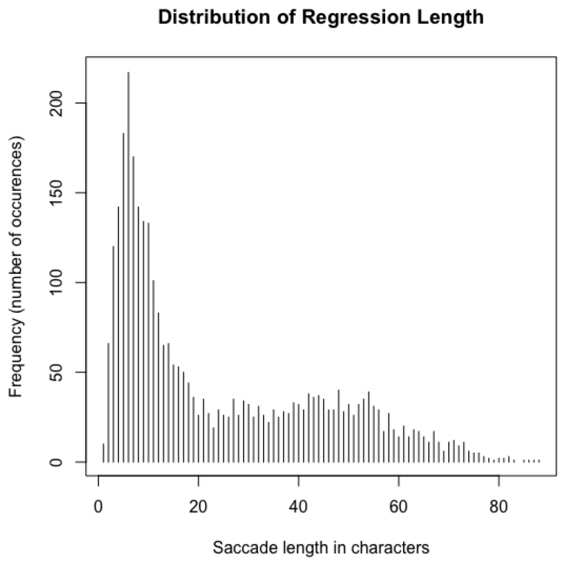

Figure 8: *Distribution of Regression Length. Shows the number of occurrences (y-axis) of all regressions that were initiated in the final region (n=3267) with the corresponding length in characters (x-axis).*

Because the length of a regressive saccade cannot be fully distinguished from the position in the sentence where it was launched, we additionally examined the landing site distribution of regressions with regard to their target position (in characters from sentence beginning). Again, we only took regressions that were initiated from the final region in order to capture a broader variety of possible target positions.

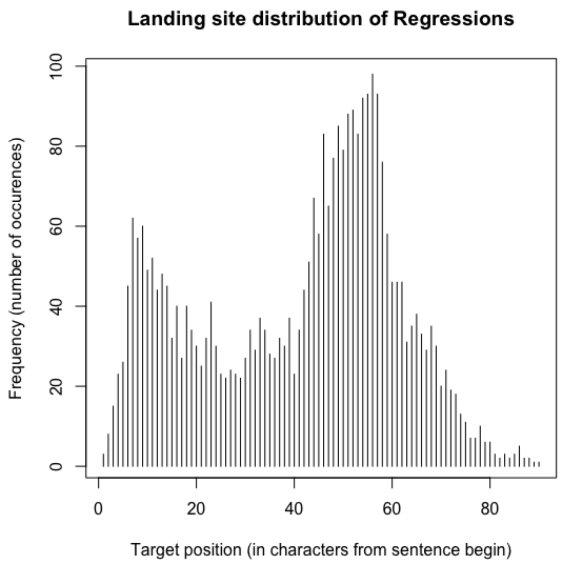

Figure 9: *Landing site distribution of Regressions. Shows the number of occurrences of regressions that were initiated in the final region for a particular position from the sentence beginning (in characters).*

The visual inspection (see Figure 9) reveals that the landing sites of regressions are not equally distributed over the whole sentence. Rather, they show two peaks: One group of the regressions targets the sentence beginning and a second group targets a position much closer to the position where they have been launched. Together with the previous examination, this pattern suggests that regressions can be assigned to two groups based on their target position: One group of regressions targets the sentence beginning whereas the second group targets a position within the 15 characters to the left of the current fixation.

In order to examine the target position of regressions in the sentence, we aligned the target position of the regressions with the sentence regions defined above. For this analysis, we took the first four regions in the sentence (initial region, subject region, verb region, object region; see Table 1) and further divided the original last region into a pre-final and a final region (consisting of the last 2-3 words of the sentence) as exemplified in Table 6, so that we had six regions in total. This allowed us to get a more fine-grained picture of the regression pattern at the end of a sentence. In addition, we excluded all regressions that landed within the same region, so that 6309 regressive saccades underwent analysis.

Since we primarily expected an effect of sentence type on the landing site distribution of regressions, we calculated the proportion of regressions that went into a particular region for each sentence type separately (see Figure 10).

The results show that there are great differences between the landing site patterns of the two sentence types: Whereas 44.2% of all inter-region regressions in anomalous sentences targeted the subject and verb region, this proportion dropped to 31.39% for non-anomalous sentences. By contrast, 34.02% of all inter-region regression in non-anomalous sentences targeted the initial region, which was only true for 27.30% of the regression in anomalous sentences. Chi squared tests showed highly significant differences between the proportions for the two sentence types in all regions except region 4 (see Figure 10)[32].

[32] It is problematic to compute chi squared tests for goodness of fit for each region separately without correcting for multiple comparisons. However, because we were interested in the difference of regression targets between sentence types, we were neither able to compare the proportion of regressions for regions (due to the unequal region size) nor to use binominal data for logistic regression models (due to the unequal number of regressive eye movements for anomalous vs. non-anomalous sentences). Thus, we decided to compute comparisons of proportions for each region separately but we acknowledge that the results have to be viewed with caution.

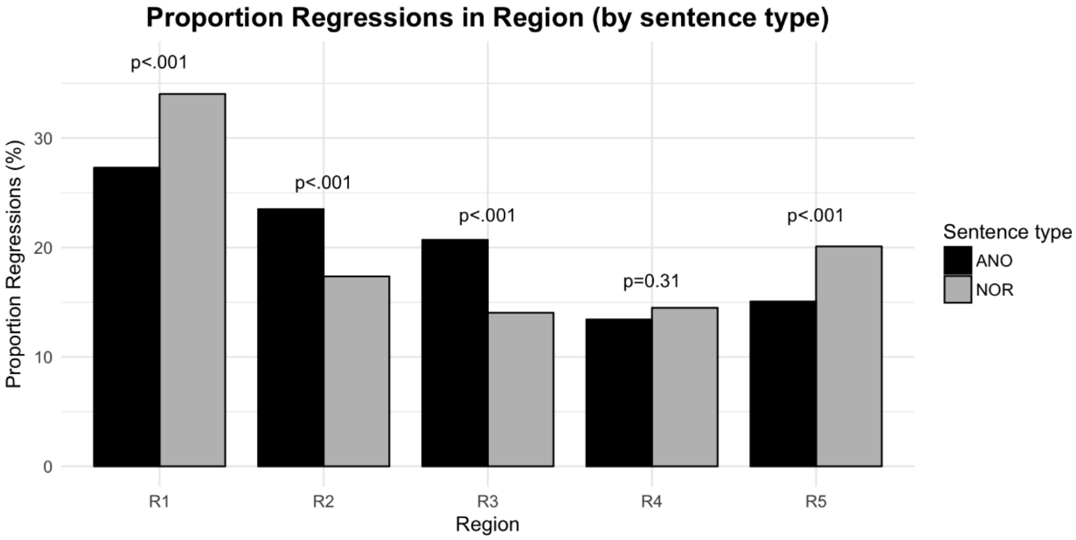

Figure 10: *Proportion of Regressions in Region by sentence type. Shows the distribution of all inter-region regressions (proportion from all inter-region regressions of a sentence type in percent) of a particular sentence type with regard to the region where the regression landed. Chi squared test compared the difference of proportions between sentence types for each region. Abbreviations: ANO = anomalous sentences, NOR = non-anomalous sentences, R1 = initial region, R2 = subject region, R3 = verb region, R4 = object region, R5 = pre-final region.*

In addition, we investigated the impact of question difficulty on the landing site pattern. The procedure was the same as for the analysis of sentence types.

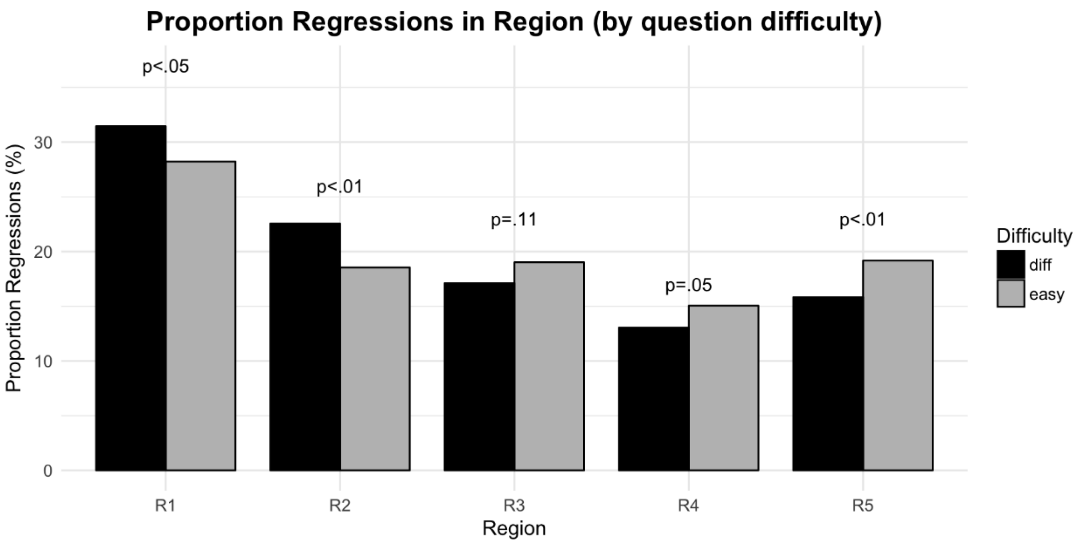

Figure 11: *Proportion of Regressions in Region by question difficulty. Shows the distribution of all inter-region regressions (proportion from all inter-region regressions of a sentence type in percent) of a particular question type with regard to the region where the regression landed. Chi squared test compared the difference of proportions between question types for each region. Abbreviations: diff = difficult questions, easy = easy questions, R1 = initial region, R2 = subject region, R3 = verb region, R4 = object region, R5 = pre-final region.*

The results (see Figure 11) show a smaller numerical influence of question difficulty on landing sites than sentence type, but chi squared tests also revealed significant differences between the two question types for region 1, 2 and 5. This indicates that in the difficult condition more regressions were targeting the sentence beginning, whereas in the easy condition the readers were more likely to regress to the immediately preceding region.

Fixation duration before regressions

The second question we addressed in the word-based analysis of regressive eye movements is whether there is a difference between the duration of the fixations before a saccade is triggered, depending on the saccade type (progressive vs. regressive).

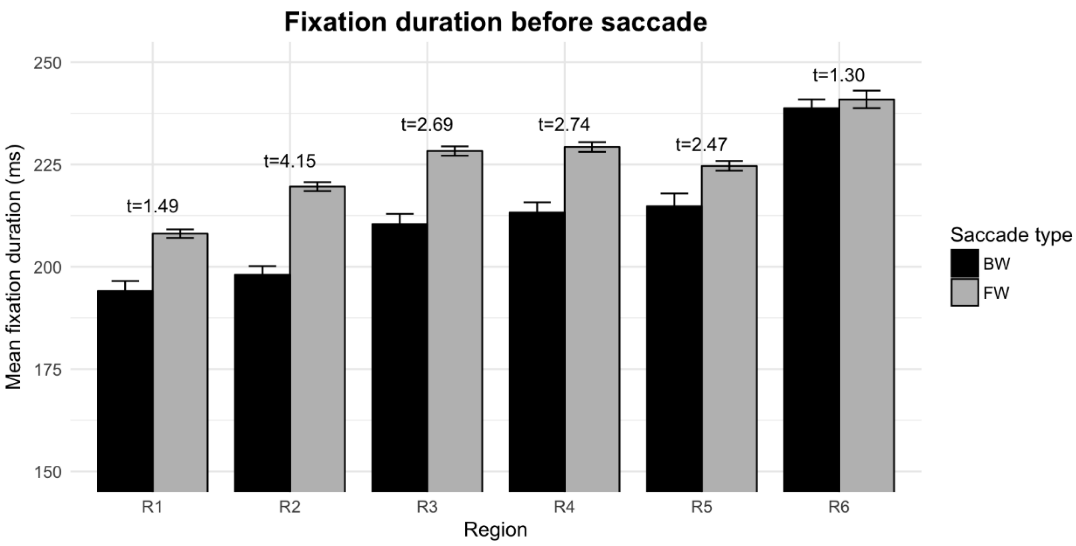

Figure 12: *Fixation duration before saccade. Mean fixation durations and standard errors are given for each saccade type separately. T-values represent the results from the linear mixed-effect models for the main effect of saccade type (see text for further information). Abbreviations: BW = regressive saccade, FW = progressive saccade, R1 = initial region, R2 = subject region, R3 = verb region, R4 = object region, R5 = pre-final region, R6 = final region.*

A first analysis revealed that fixations before regressions were generally shorter (mean 217.06ms) than fixations before progressive saccades (mean 222.84ms). This difference was highly significant ($t(14691) = 4.92$, $p<0.001$). In order to see whether there was an influence of sentence type, question difficulty or a difference between the different regions in the sentence, we calculated linear mixed-effects models of log fixation duration for each of the six regions separately (see the details of the regioning scheme above). These models included anomaly, difficulty and sentence type and their interaction as fixed effects as well as random intercepts for subjects and items with the maximal random effect structure (difficulty was only included in the by-item slopes). The results showed that sentence type was significant in the subject (β = .17, SE = .04, t = 4.15), verb (β = .08, SE = .03, t = 2.69), object (β = .09, SE = .03, t = 2.74) and pre-final region (β = .10, SE = .04, t = 2.47) but not in the initial (β = .06, SE = .04, t = 1.49) and final region (β = .05, SE = .04, t = 1.30; see Figure 11). No other effects or interactions became significant. In sum, fixation durations were significantly shorter before regressive saccades compared to progressive saccades, but not at the end of a sentence.

5.3 Discussion

The present experiment investigated how implicit modulations of reading strategy may influence eye movements during sentence reading. We manipulated the difficulty of comprehension questions between subjects who read sentences containing different types of anomalous and non-anomalous material. The results are easily summarized: The difficulty of the comprehension questions did not significantly affect first-pass reading of sentences; instead, participants were more likely to reread the sentence after regressing from near the end, and/or take longer in rereading, when faced with difficult comprehension questions. Importantly, this pattern was replicated for all three sentence types in our study, and held for sentences with and without semantic reversal anomalies (SRAs). Indeed, the task manipulation did not influence readers' eye movement behavior until they reached the final two or three words of the sentence.

In accordance with our experimental goals we will now first discuss these results with regard to the predictions of the *Information Gathering Framework*, before we turn to a detailed discussion of their implications for the language processing mechanisms.

The *Information Gathering Framework* that we introduced in the beginning of this thesis makes several predictions for the behavior of regressive as well as progressive eye movements during sentence reading. Some of these predictions were explicitly tested in this experiment, whereas others were examined by post-hoc analyses. We will travers through the single findings and discuss their relevance for the *Information Gathering Framework* in more detail.

First, the effect of task on certain aspects of the eye movement behavior suggests that we were indeed able to induce different thresholds for the execution of saccades by implicitly modulating the difficulty of comprehension questions. Importantly, because the question manipulation only affected the probability of regressions and not the global reading pattern (like first pass reading), we could conclude that we were able to adjust the backward threshold separately from the forward threshold. Thus, tasks that focus on sentence comprehension and that require a processing of the linguistic input on all linguistics levels shape the execution of regressive but not of progressive saccades which fits well with the assumptions of the *Information Gathering Framework*.

Second, the *Information Gathering Framework* makes the strong prediction that question difficulty should modulate the likelihood for regressions due to missing evidence but not due to integration difficulties. Our results from the SRA sentences confirm this hypothesis: Question difficulty did increase the probability to regress from the end of a sentence (which is likely to reflect missing evidence) but surprisingly it did not affect the probability of regressions in the sentence areas where the anomaly became apparent (verb and object region). Our data therefore presents evidence for the first time that regressions can be distinguished into two groups depending on their cause (regressions due to integration difficulties and due to missing evidence) and that these regressions behave separately in terms of their sensitivity to higher order task manipulations.

Third, an inspection of the landing site patterns of regressive eye movements by sentence type revealed that the target position of regressions was not subject to a random mechanism but was rather determined by the linguistic task. Thus, regressions in the anomalous sentences were more often targeting the subject and verb region than regressions in non-anomalous sentences. Given that these regions contain a potential source for the interpretation difficulties, they seem to provide an appropriate location for targeted repair (despite the fact that repair was not possible in the anomalous sentences). In the non-anomalous sentences, by contrast, where there seems no need for a targeted repair of the adopted sentence structure, regressions were more likely to target the sentence beginning or the immediately preceding region. This could be part of a regression strategy because this strategy needs less linguistic knowledge and is not determined by a certain linguistic problem.

Surprisingly, the question difficulty had only a small effect on the landing positions of regressions. However, readers faced with more difficult comprehension questions were more likely to regress to the sentence beginning, whereas readers in the easy condition were more likely to regress to the immediately preceding region. One reason for this behavior could be that rereading from the sentence beginning takes more time than just rereading the preceding region so that readers in the easy condition tried to avoid this time-consuming process (in accordance with the speed-accuracy trade-off). A more plausible explanation, however, is given by the architecture of the *Information Gathering Framework* itself: If the manipulation of question difficulty only affects the backward threshold and if this threshold in turn only affects the execution of regressions in response to missing evidence (as our data suggests), we would expect that these regressions do not target a particular, strongly linguistically determined sentence region because no repair is necessary. Instead, we would expect that these target positions are more likely to be guided by a general checking strategy.

Fourth, the *Information Gathering Framework* assumes that the regression target is selected on the basis of linguistic computations only within a limited window of attention, comprising about 15 letters to the left of the current fixation. Although our experiment was not designed to test this hypothesis, we were nonetheless able to find some hints for the reliability of this assumption. On the one hand, our results show that the majority of regressions were of 15 or less characters in length. Although it was descriptive evidence only, it seems that there was some kind of invisible boundary at this point because the number of regressions with a length of 16 characters or more decreased clearly but was almost stable for the remaining characters. On the other hand, the target positions from the sentence beginning indicate that regressions can be distinguished into two groups based on their target position: One group of regressions targeted the sentence beginning and the other group a position that was much closer to the current fixation. Since the region where the anomaly became apparent (verb and object region) was often within the 15-character window when regressing from the final region, we cannot fully clarify if these shorter regressions were made in response to a linguistic computation (as the increased number of regressions in the subject and verb region for anomalous sentences suggests, c.f. Figure 10) or due to a strategy that targeted the preceding word / region. Thus, we have to leave for future research to determine exactly the role of the perceptual span in regression target selection. Nonetheless, the distinction between two regression target patterns is perfectly in line with the *Information Gathering Framework* that distinguishes between a strategy-based and a linguistically-based target selection.

Fifth, the examination of fixation durations before regressive and progressive saccades, respectively, revealed shorter fixation durations before regressions compared to progressions. Whereas these findings replicate earlier findings (Altmann et al., 1992), our analysis also revealed that these shorter fixation durations were not found at the end of the sentence. Whereas none of the models proposed in the literature may account for these findings which are on the first glance counterintuitive, our *Information Gathering Framework* explicitly predicts such a behavior: Because regressions in response to integration difficulties are triggered whenever the confidence level of the particular word n falls under the forward threshold (during a fixation on word $n+1$), this happens before the confidence level of word $n+1$ reaches the forward threshold. Thus, shorter fixation durations are expected before regressions in response to integration difficulties (c.f. Figure 5). For regressions that are triggered in response to missing evidence, by contrast, the confidence level just does not reach the backward threshold within the time it takes for the confidence level of word *n+1* to reach the forward threshold. Thus, the fixation duration should not be shortened before making a regression due to missing evidence. In our experiment, we were able to distinguish these two groups because regressions at the end of a sentence should primarily reflect a response to missing evidence. Thus, the lack of a difference between fixation durations for the two types of eye movements (progressive vs. regressive) at the end of a sentence is perfectly in line with the assumptions of the *Information Gathering Framework*.

In sum, our data provide great support for the *Information Gathering Framework*. Although not all results have the capability to question the architecture of the model fundamentally, they nonetheless revealed clear evidence for a variety of predictions of the *Information Gathering Framework*.

In the second part of the Discussion, we will now turn to the implications of the results for the language processing system by relating them to current debates in the sentence processing literature.

In our study, we were able to extend WK's findings in several ways. Firstly (as already discussed above), our results confirm, with a more pronounced question difficulty manipulation, WK's finding that such a manipulation primarily modulates reading behavior by influencing the probability of rereading, at least with younger adults. Second, while WK showed that regressions were more common when readers were faced with difficult questions, their word-level analyses did not investigate the point in sentences from which these regressions were launched; our data clearly show that question difficulty influenced the subject's decision to reread only when he or she reached the final words of the sentence.

Finally, we tested whether comprehension demands have a particular impact on reading of anomalous sentences, and we found that they did not: Anomaly effects during incremental processing were as pronounced with easy questions as with difficult questions. Thus, the results do not support the idea that comprehension demands determine whether readers engage in syntactically-licensed assignment of thematic roles. A clear limitation is that we do not have direct evidence about the final interpretations of SRA sentences, because we did not include comprehension questions for these sentences; it is possible that readers may have sometimes misunderstood these sentences, and that this was more common with easy comprehension questions. However, the on-line effect of anomaly was not contingent on comprehension demands. We acknowledge that to some extent this issue may depend on whether sentences involve canonical (in this case, Subject-Verb-Object) word order, or a non-canonical order such as in passive or relative clause structures. Previous studies (e.g., Christianson, Luke, & Ferreira, 2010; Ferreira, 2003) have found that comprehenders are most likely to misinterpret thematic roles when a sentence involves non-canonical word order. Thus, it is possible that comprehension demands would have a clearer influence on incremental thematic role assignment with non-canonical word order.

Our results from sentences with SRAs do also have implications for the functional architecture of the language processing system. Though the processing of these structures was not affected by task condition, processing was influenced by the lexical association between subject and verb. The anomaly effect on reading times on the verb was reduced when the subject was a highly plausible theme for the verb (*the flower is picking* vs. *the flower is drawing*). On the object, by contrast, this effect was reversed, with a larger anomaly effect in the associated conditions. On its surface, this pattern suggests a very slight delay in the detection of a semantic anomaly when the subject is highly associated with the verb, and is a likely theme. It is important to note, however, that a sizable anomaly effect was still present at the verb even when the subject and verb were highly associated: about 30 ms in first pass time (averaging across task conditions) and about 60 ms in go-past time. Thus, the present results do not directly support the claim from the ERP literature (Kim & Osterhout, 2005) put forward to account for "semantic P600" effects in the absence of N400 modulations, namely that readers initially assign thematic roles in SRA sentences based on plausibility, rather than based on the actual syntactic structure. Our results may therefore be regarded as consistent with more recent ERP studies (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Stroud & Phillips, 2012; Brouwer, Fitz & Hoeks, 2012) that emphasize the role of lexical pre-activation in explaining the lack of N400 amplitude modulation by SRAs in languages such as English (note that across languages, the situation is more complex; cf. Bornkessel-Schlesewsky et al., 2011; Tune et al., 2014). The present results demonstrate that in reading, the anomaly effect is present at the verb in both association conditions, but is reduced when lexical association between subject and verb is high.

Finally, our data also adds some more insights into the role of regressive eye movements in sentence processing. As already mentioned in the first part of the Discussion, regressions are modulated by the sentence type used in the experiment. Thus, the probability to regress increased in the verb, object and final region in anomalous sentences, a finding which has also been reported by other eye-tracking studies (e.g., Rayner et al., 2004). According to the *Information Gathering Framework*, these regressions serve to confirm that the sentence is indeed anomalous by increasing the lexical quality level of the anomalous material (also indicated by the increased likelihood to regress into the subject and verb region for anomalous sentences as evidence for targeted repair); and if so, a regression might lead to an increased probability of correctly recognizing that the sentence is anomalous. However, in the absence of explicit comprehension probes of anomalous sentences (which have also been absent from other eye-tracking studies investigating processing of implausible or anomalous sentences, e.g., Rayner et al., 2004), this cannot be confirmed.

In addition, as for the results reported by Schotter and colleagues and discussed in the first part of this thesis (Schotter et al., 2014), our data also failed to show evidence of improved comprehension when readers regressed from the end of the sentence (see also Christianson, Luke, Hussey, & Wochna, in press, for a similar finding). Schotter et al. interpret their null finding as suggesting that regressions may be compensatory, improving comprehension up to, but not beyond, the level of comprehension in cases where the reader does not feel the need to reread. We think this is a very plausible interpretation. The overall improvement in comprehension when readers have the opportunity to regress, coupled with the lack of comprehension benefit when they actually do regress, is expected if there are some cases in which first-pass reading yields an incomplete or defective representation of sentence meaning, and regressions are selectively initiated on these trials in order to improve comprehension. But it is also expected if the control of regressive eye movements depends on the linguistic knowledge of the reader as for example proposed by the *Information Gathering Framework*. Thus, the reader may trigger a regression in response to integration difficulties, but in order to solve these difficulties (in our case to correctly reanalyze a garden path sentence or an object relative clause structure) he requires a detailed analysis of the linguistic input and a knowledge of the possible sentence structures in English. In sum, it is very likely that some readers did not understand the sentence correctly, even if they would have read the sentence several times because they were just not able to understand the sentence due to a lack of linguistic knowledge.

However, our data also indicates that regressions are made in response to sentence processing tasks others than a repair of integration failure. We provided various evidence that regressions from the end of a sentence behave differently from those initiated in response to anomalous sentence structures. The *Information Gathering Framework* associates them with missing evidence. In our case, the end of a sentence was (as usually) indicated by a full stop, but also, the end of a sentence required a motoric task by pressing a button and the sentence disappeared from the screen (which may have additionally increased the number of end-of-sentence regressions). This may have made the end of a sentence a very prominent area because it completes a unit and enables the reader to evaluate the information within this unit. Accordingly, it has been proposed in the literature that at the end of each sentence a sentence “wrap-up” mechanism (Hirotani, Frazier, & Rayner, 2006; Rayner, Kambe, & Duffy, 2000) is deployed. This assumption receives more support from the finding in our data that fixation durations in general tended to be longer at the end of a sentence. This could be due to the matching of all syntactic and semantic predictions of the sentence (like open syntactic nodes, predictions about thematic roles and so forth) where nothing can be postponed to the next sentence instead of just integrating the current input into the sentence structure. Also, our data suggest that at the end of a sentence the confidence into the meaning of the whole sentence is evaluated and matched with the requirements of the comprehension question. If this evaluation yields that the expected supporting evidence for the current sentence interpretation is not provided by the current input, the reader regresses in order to gather additional information (depending on the amount of information that is required in order to answer the comprehension questions). Probably these kinds of regressions are primarily triggered in order to confirm an established sentence interpretation and to refresh the lexical representations and not in order to find a solution for an integration problem. But this hypothesis has to be addressed by future research.

5.4 Summary

In the present experiment we investigated the influence of different question types on the eye movement patterns when reading anomalous and non-anomalous sentences. The results provided clear evidence that the task manipulation did not affect first pass reading but did only increase the probability to regress from the sentence final region. Importantly, this pattern was replicated for all three sentence types used in the experiment, with and without an anomaly. In addition, we examined the pattern of regressions in more detail and were able to show that regressions initiated from the sentence final region behaved differently to those triggered as a response to integration difficulties (e.g., in terms of preceding fixation durations).

We interpreted these findings as clear evidence for the *Information Gathering Framework*. The results can be best explained and are directly predicted by a model that assumes that two different scenarios exist that cause a regressive eye movement. The resulting groups of regressive eye movements are hypothesized to be differently affected by the question and anomaly manipulation which has been confirmed by our results.

In addition, our data has also methodological and linguistic implications, showing that different comprehension questions were not able to eliminate the detection of anomalies, although we used sentence structures that were constructed in order to increase the likelihood to overlook these anomalies (so called Semantic Reversal Anomalies). This suggests that language processing is always complex and comprises an analysis on higher linguistic levels like syntax and semantics as well.

The experiment also points to an important field for future research: the comparison of different regression types. Whereas all regressions can be subsumed under a general function which is to gather additional information relevant in the course of sentence processing, they can be further distinguished into two groups. Especially end-of-sentence regressions have not been in the broader research focus so far and more evidence is needed to determine their exact role in sentence processing. Thus, we think that experimental designs that allow to examine these two types of regression in more detail, provide important insights into the control of regressive eye movements during reading in general.

6. Experiment 2

6.1 Introduction

In Experiment 1 we have examined the impact of question difficulties on regression behavior during sentence reading while the eyes of the participants were tracked. The results overall suggest that the execution and target selection of regressive eye movements is highly linguistically constrained which also fits well with the assumptions of the *Information Gathering Framework*. However, whereas eye tracking experiments provide a very helpful tool in exploring the role of regressive eye movements in reading in general, the measurement of reading times and regression rates tells us little or nothing about the neural architecture of reading in particular, which is one of the most challenging goals for current research (Reichle, Tokowicz, Liu, & Perfetti, 2011).

Specifically, most reading models aim to reflect mental processes by using (at least implicitly) concepts such as attention or cognitive control. But these concepts can hardly be examined on the basis of reading patterns only. Also, since mental processes are thought to be carried out by the brain, the proposed reading models should be plausible from a neural perspective. Thus, the mental processes proposed by the reading models should additionally correspond to the neural architecture of the brain, if we really aiming to model human reading behavior.

An important prerequisite for aligning current reading models with the neural architecture of reading, however, is to examine the neural correlates of reading by using high-spatial resolution techniques such as functional magnetic resonance imaging (fMRI). In the last decades, this method has been deployed to study a wide range of cognitive tasks like memory or language processing. But in the context of reading, only artificial designs such as rapid serial visual presentation (RSVP) experiments were used, primarily due to methodological requirements of the fMRI that were thought not to fit to natural reading (e.g., the reader-dependent variability in reading behavior; see Marsman, Renken, Velichkovsky, Hooymans, & Cornelissen, 2012, for a discussion). Thus, the neural correlates of natural reading still remain largely unknown.

Recently, however, experimental settings with co-registering eye-tracking / fMRI paradigms were introduced that allow for an investigation of the neural correlates of free reading by modelling fixation onsets as events for the blood oxygenation level dependent (BOLD) response (Marsman et al., 2012; Richlan et al., 2014; Henderson, Choi, Luke, &

Desai, 2015). This, in particular, makes it possible to explore several important features of reading that occur only in free reading conditions; for example, the execution of saccadic eye movements or fixation durations.

The latter feature was in the focus of a study conducted by Henderson and colleagues (Henderson et al., 2015), who investigated the neural correlates of fixation durations during text and pseudo-text reading. They found that regions of visual and attentional processing (e.g., medial superior frontal gyrus [SFG], including supplementary eye fields [SEF] and supplementary motor area [SMA]) as well as the ventral language network (e.g., medial and superior temporal gyri [MTG / STG], superior temporal sulcus [STS]) showed increased activations for fixation durations during text compared to pseudo-text reading. In particular, the increased activation in the SEF was interpreted by the authors as a potential component of the control network that is implemented into current models of eye movement control such as SWIFT and E-Z Reader. By contrast, negatively correlated activations in the supramarginal gyrus (SMG) were assumed to reflect lexical control of fixation duration as also proposed by E-Z Reader.

Since fixation durations reflect processing difficulties, they represent a perfect starting point for exploring the neural correlates of natural reading. However, a second important component of reading that is assumed to be sensitive to linguistic processing is regressive eye movements. It is therefore logical to extend the exploration of natural reading to the investigation of regressive eye movements.

The neural underpinnings of eye movements in general have been in the broader focus of research since several years (see e.g. Schutz, Braun, & Gegenfurtner, 2011, for a recent review on eye movements and perception). But although eye movements occur frequently in all visual tasks as in scene perception or visual search, their role in reading is somewhat special. The pure visual input in sentence or text reading, especially when using Western European writing systems, usually allows for several predictions that may constrain the execution of eye movement sequences[33]. First, the visual input is arranged on a horizontal line and has an intended reading direction, namely from left to right. Secondly, the visual input is permanent, thus allowing for parafoveal input to be used for

[33] For the current purpose, we will only discuss the characteristics of Western European writing systems. Other writing systems share the similar characteristics but may differ with regard to some details (e.g., the reading direction). In addition, it has to be noted that all of these principles may be violated in some special situations (e.g., for handwriting).

saccade planning. Thirdly, because the visual input is formatted consistently, the characteristics of the current visual input allow for predictions about the following visual input, such as letter size, line length, font type, and so forth.

These considerations show that it is plausible to assume that saccade planning in reading may be mainly driven by these visual features and that the eyes move through the sentence based on an automatic scanning mechanism. Accordingly, it has been proposed in the literature that the decision as to when and where to move the eyes may be subject to an oculomotor strategy that works largely independent of cognitive and linguistic control (O'Regan, 1990, 1992; O'Regan & Lévy-Schoen, 1987; Yang & McConkie, 2001, 2004; Yang, 2006, for a more recent review see Vitu, 2011). Yang and McConkie (2001, 2004), for example, assume in their competition / interaction theory of eye movement control during reading that only late saccades (i.e., saccades following rather long fixation durations) could be influenced by cognitive control, whereas earlier saccades are triggered by a default execution mechanism that is not sensitive to input from the currently-viewed text. Further support for this claim derives from the consideration that the time window of a fixation during reading (about 200–250ms) is usually too short to affect the planning of the subsequent saccade.

Reading, however, is not a procedure that simply aims for the intake of visual input. Rather, it aims for the decoding of linguistic information. Thus, the opposite claim has also been strongly brought forward, namely, that eye movement generation in reading is highly influenced by cognitive and linguistic factors. Bicknell and Levy (2012) for example argue that word identification is the major determinant of eye movement control, thus sending the eyes to the position in the text that allows for optimal and efficient identification of the following words. Thus, saccade programming is assumed to be highly coupled with linguistic processing and to be subject to cognitive control (for a more recent review see Rayner & Liversedge, 2011).

Whereas these factors have been primarily discussed in the context of progressive saccades, they can also be applied to regressive eye movements during reading (see chapter 2.1 for a detailed discussion). It is here that the question of the strength of attentional or linguistic control monitoring the execution and target selection has been asked. Importantly, whereas this issue has been (to our knowledge) examined only within behavioral paradigms so far, the engagement of attentional or linguistic control in the two types of eye movements (regressive vs. progressive) should also be reflected in the neural activation patterns these saccades elicit.

The *Information Gathering Framework*, in particular, makes some clear predictions regarding the differences between progressive and regressive saccades respectively, that are expected to be mirrored in the neural response (albeit that the framework does not explicitly model neural correlates of reading). Let us shortly consider these differences.

Whereas the forward mechanism is assumed to work in a highly automatic manner, per default targeting the next word (although the forward mechanism is sensitive to linguistic processing since word identification is its primary task), a regressive saccade interrupts this forward walk. This interruption is caused by a prediction error, either because a prediction of the production rules is violated or because expected evidence is missing. In response, a regressive eye movement is triggered which is accompanied by a shift of attention to the left of the current fixation. Also, a re-computing of the confidence levels within the perceptual span takes place and a target is selected, either on the basis of this linguistic process or on the basis of a strategy.

If the assumptions of the *Information Gathering Framework* are correct, we should expect that a regressive eye movement is much more than just an eye movement in the opposite direction, and that the qualitative differences between regressive and progressive saccades should also lead to different neural activation patterns. In particular, we hypothesize that if regressions are correlated with prediction error and lead to an attention shift to the left, they should engender increased activations in regions associated with attention control (e.g. inferior parietal sulcus [IPS]) and conflict resolution (e.g. anterior part of the cingulate gyrus and sulcus [ACC]) compared to progressive saccades. Due to the absence of a default mechanisms, we would also expect that the generation of regressive eye movements (i.e. the target selection) requires a higher engagement of regions associated with eye movement control (e.g., medial superior frontal gyrus [SFG]) compared to the more automatic generation of progressive eye movements.

To examine these predictions, we conducted a concurrent fMRI / eye-tracking experiment of subjects reading anomalous and non-anomalous German sentences (experiment 2b). We followed the procedure of Henderson et al. (2015) but instead of modelling fixation onsets we related the hemodynamic response to the onsets of saccades, for each saccade type separately. However, because the *Information Gathering Framework* is not specified with regard to second pass reading as yet, we only took progressive inter-word saccades that occurred in first pass time. In addition, we included the length of the saccade as a parametric regressor which allowed us to investigate the neural activations correlated with increased saccade length. This is of particular interest in the context of *the Information Gathering Framework* that predicts different target selection mechanisms for regressions that should also be correlated to saccade length.

Since it is still unclear whether the eye movement pattern of participants lying in an fMRI scanner can be compared with the reading behavior of people in typical eye-tracking settings outside the scanner, we further tested these differences by deploying the same design in a single eye-tracking experiment (Experiment 2a) and a combined fMRI / eyetracking experiment (Experiment 2b). Given that Experiment 1 revealed a high impact of question difficulties on regression behavior at the end of a sentence, we might view the experimental settings (single vs. combined) as different tasks and predict that this should also affect regression behavior. The results will tell us if this was the case or not.

In addition, the experiment was designed from a linguistic perspective by using German counterparts of the English SRA sentences employed in Experiment 1 in order to test for crosslinguistic differences in the processing of these structures. In particular, because German and English differ with regard to their mapping of linguistic functions (like the actor role) with forms (like word order; see e.g. MacWhinney et al., 1984), we might expect that information cues like the animacy of a particular argument are used differently when sentences with a reversal of thematic roles are read. In particular, native speakers of German should rely mostly on case marking information, whereas in English word order is proposed to be the most important cue to actor identification. By comparing the reading behavior of the participants in Experiment 1 with the reading behavior of the participants in the two Experiments 2a and 2b, we expect to find these differences.

6.2 Method

Since the study consisted of two experiments, the single eye-tracking experiment (Experiment 2a) and the combined fMRI / eye-tracking experiment (Experiment 2b) which required different prerequisites and technical settings, we will report the methodological specifications separately when they differed between the two experiments.

6.2.1 Participants

6.2.1.1 Single eye-tracking experiment (Experiment 2a)

Twenty-five healthy subjects participated in Experiment 2a (22 female, mean age 23.96 years, range 19–38) who were either students from University Marburg, Germany, or people from Marburg area. All of them were monolingually raised, native speakers of German and had normal or corrected-to-normal vision. All participants were naïve concerning the purpose of the experiment and received about 10€ for their participation.

6.2.1.2 Combined fMRI / eye-tracking experiment (Experiment 2b)

Twenty-two healthy students (13 female, mean age 23.63 years, range 19–31) of the University of Marburg, Germany participated in Experiment 2b. One additional subject was excluded due to unusably long reading times that did not allow for the experiment to be completed. All participants were right-handed (indicated by an adapted German version of the Edinburgh Handedness Inventory; Oldfield, 1971), monolingually raised native speakers of German and reported neither impairments with regard to reading and writing ability nor neurological, psychiatric or other disorders. All participants met the safety and participation requirements and gave written informed consent prior to participation. They were naïve concerning the purpose of the experiment and received 20€ for their participation. The study procedure conformed to the Declaration of Helsinki and was approved by the ethics committee of the Medical Faculty at the University of Marburg. None of the subjects had participated in the single eye-tracking experiment.

6.2.2 Materials

The sentence material used here was the same for both Experiments 2a and 2b. Since we were (besides the neural underpinnings of regressive vs. progressive saccades) also interested in the interaction of syntax and semantic processing during sentence reading, we created German-counterparts of the English SRA sentences used in Experiment 1. This was done to allow for a direct comparison of Experiment 1 and 2 in two different languages and to have an additional testing tool, especially in view of the reliability of the combined fMRT / eye-tracking experiment.

Table 7*: Example stimuli of the German SRA sentences used in Experiment 2. Sentence regions for the region-based eye-tracking analysis are marked by the vertical strokes.*

Semantic Reversal Anomalies (SRA) – German		
a)	**Der Lehrer | korrigiert | den Fehler | nach zwanzig Berufsjahren schon | fast automatisch.** *The teacher$_{(NOM)}$ corrects the mistake$_{(ACC)}$ after twenty years of experience almost automatically.*	*non anomalous, high associated*
b)	**Der Lehrer | errät | den Fehler | nach zwanzig Berufsjahren schon | fast automatisch.** *The teacher$_{(NOM)}$ guesses the mistake$_{(ACC)}$ after twenty years of experience almost automatically.*	*non anomalous, low associated*
c)	**Der Fehler | korrigiert | den Lehrer | nach zwanzig Berufsjahren schon | fast automatisch.** *The mistake$_{(NOM)}$ corrects the teacher$_{(ACC)}$ after twenty years of experience almost automatically.*	*anomalous, high associated*
d)	**Der Lehrer | errät | den Fehler | nach zwanzig Berufsjahren schon | fast automatisch.** *The mistake$_{(NOM)}$ guesses the teacher$_{(ACC)}$ after twenty years of experience almost automatically.*	*anomalous, low associated*

An example of the SRA sentences employed is given in Table 7. Because it is not possible in German that two constituents precede the main verb in standard declarative sentences (in contrast to the English sentence structures), in all of our SRA sentences only one argument (which always was the subject) preceded the verb. Although an object-first structure is possible in German, all arguments were unambiguously case-marked on the masculine determiner “der” (nominative) or “den” (accusative), respectively, so that they could be unequivocally identified as the subject or object, respectively.

As in the English counterparts, one of the two arguments was animate (e.g. “Lehrer” [*teacher*]) and the other was inanimate (e.g., “Fehler” [*mistake*]). When the inanimate argument was used as the subject, this resulted in implausibility (“Der Fehler korrigiert den Lehrer” [*the mistake corrects the teacher*]). In addition, we manipulated the degree of association between the verb and the inanimate object (high: “Fehler – korrigieren” [*mistake – to correct*] vs. low: “Fehler – erraten” [*mistake – to guess*]), leading to a 2 x 2 factorial design (ANOMALY x ASSOCIATION). In total, we created 72 experimental stimulus sets of four stimulus versions each (all stimulus sentences can be found in the Appendix).

In an online pre-test, we asked a separate group of participants to rate the association between pairs of the inanimate noun and verb on a 5-point scale (1 – very high associated, 5 – not associated) and the acceptability of the sentences, again on a 5-point scale (1 – not acceptable, 5 – very acceptable). The word pairs and sentences were assigned to four lists separately and each list was rated by 18 people (for the association judgement) and 20 people (for the acceptability judgement). The analysis revealed that the word pairs in the low associated condition were significantly lower associated than in the high associated condition (mean verbs low: 3.70, mean verbs high: 1.44; $t(74.40) = -24.93$, $p<.001$). Also, the non-anomalous sentences were significantly more acceptable than their anomalous counterparts (mean anomalous sentences: 1.27, mean non-anomalous sentences: 4.49, $t(202) = 57.93$, $p<.001$).

There were no differences between the word length of the high and low associated verbs (mean verbs high: 6.49 characters, mean verbs low: 6.71 characters; $t(141.76) = -0.71$, $p = .48$) or between the animate and inanimate nouns (mean animate nouns: 6.83 characters, mean inanimate nouns: 6.42 characters; $t(136.77)=1.20$, $p = .23$), neither were there differences between the frequency of the two verb types (frequency based on the frequency classes provided by the Wortschatz Leipzig; mean verbs high: 11.31, mean verbs low: 11.21; $t(140.96) = 0.21$, $p = .83$) or noun types (mean animate nouns: 11.72, mean inanimate nouns: 11.29; $t(140.55) = 1.16$, $p = .25$).

The sentences of the 72 stimulus sets were allocated to 4 experimental lists according to a Latin Square design so that each subject read 72 SRA sentences. In addition, we created 144 filler sentences, containing anomalous and non-anomalous sentence material, that were used to prevent the readers from developing certain reading strategies. Additionally, some of these sentences were constructed for another experimental manipulation which is of no interest for the present purpose. 40 of the filler sentences were followed by a simple YES / NO comprehension question to ensure that the subject read the sentences carefully. The 72 stimulus sentences were randomly intermixed with the 144 filler sentences, so that each subject read 216 sentences in total (60 anomalous). All sentences were syntactically well-formed German sentences.

6.2.3 Procedure

6.2.3.1 Single eye-tracking experiment (Experiment 2a)

After reading the instructions and giving written consent, the participants were seated in front of a computer screen. An SR Research EyeLink 1000 tower mount tracker with a sampling rate of 1000Hz was used to record the participants' eye movements during reading. Viewing was binocular but only the right eye was tracked. All sentences were displayed on one line in 20 point Monaco font on a 22'' monitor with a refreshing rate of 60Hz that was positioned 80 cm from the participant.

Participants were instructed to read as normally as possible, "as they would read the newspaper" and were also told that some of the items might be "a little weird". Comprehension questions were presented on the screen and participants responded to them by pressing one of two buttons on a gamepad. Before the experiment began, a 9 point calibration and a short practice session of 6 practice trials was performed. The experiment took about 40 minutes in total.

6.2.3.2 Combined fMRI / eye-tracking experiment (Experiment 2b)

Before entering the scanner, participants received written instructions and performed a short practice session in which they read six sentences on a computer screen to become familiar with the procedure.

In the scanner, participants read 216 single sentences on a 92 x 52 cm computer screen (screen diagonal: 42°; mirror-to-screen distance: 203 cm) positioned at the end of the bore and reflected by a mirror mounted to the head coil. The sentences were presented in black on a grey screen in mono-spaced, size 18 font DejaVu Sans Mono. All stimuli were presented sentence by sentence on a single line. Participants responded to comprehension questions with their left hand using a button box strapped to their thigh. We chose the left hand to avoid motor-related changes in neural activity in left hemisphere networks which would be also engaged in language processing.

Each trial started with a fixation cross on the left side of the computer screen to encourage natural reading behavior. After a jittered interval (ranging from 3000 to 4000 ms, mean 3531 ms, sd 559 ms), a single sentence appeared on the screen. Participants indicated via button press that they had read the sentence. Sentence presentation was then followed by either a comprehension question or a fixation cross signaling the start of the next trial.

The stimuli were divided into four experimental blocks with short breaks in between. On average, each block lasted 8 minutes, but the duration depended on the individual reading speed of the participants (ranging from 6 to 10 minutes). Participants were instructed to read the sentences as naturally as possible. The experiment was implemented in Presentation (Neurobehavioral Systems, San Francisco, CA).

Eye-tracking Data Acquisition

We tracked eye movements during reading from the right eye using an SR Research Eye-Link 1000 Plus system with a sampling rate of 1000 Hz. Prior to the experiment, we used a horizontal 3-point calibration procedure that covered the entire presentation field. If the validation exceeded a deviation of 0.5 degrees, the calibration was repeated.

fMRI Data Acquisition

Neuroimaging was performed on a 3 Tesla MRI system (Trio, A Tim System, Siemens Medical Solutions, Erlangen, Germany) with a 12 channel head matrix receive coil. Functional images were acquired using a T2-weighted single shot echo planar imaging (EPI) sequence (TR = 1450 ms, TE = 25 ms, slice thickness = 4 mm, flip angle = 90°, in plane-resolution = 3.5 x 3.5 mm, 0.6 mm gap, FOV = 240 mm, matrix dimension = 64 x 64, 30 descending slices). The first five volumes of the functional sequence were discarded to account for saturation and stabilization effects. In addition, a whole brain T1-weighted image for each participant was collected (TR = 1900 ms, TE = 2.52 ms, slice thickness = 1mm, flip angle = 9°, matrix dimension = 256 x 256, 176 slices).

6.2.4 Analysis

Because our experiment was focusing on two different research questions, two separate analyses were performed. The first analysis was restricted to the eye movement data of the SRA sentences only and was closely aligned with the one conducted in Experiment 1. The second analysis focused on the neuroimaging data and aimed to provide insights into the neural underpinnings of progressive vs. regressive eye movements in general, why it involved the data from all sentences used in Experiment 2b.

6.2.4.1 Analysis of the reading data

Since the sentence material and the reading procedure were the same in both Experiments 2a and 2b, the eye movement data of both experiments were analyzed in exactly the same way and the two data frames were merged into one data frame (but including experiment type as a between-subjects factor). As for Experiment 1, we conducted two different analyses: a region-based and a word-based analysis.

Region-based eye movement analysis

For the region-based analysis, we deleted all trials with first-pass blinks or track losses in the verb region. Because some participants lost more than 10 trials, we excluded five subjects from Experiment 2a and one subject from Experiment 2b, so that the eye movement data of 41 participants underwent analysis (20 from Experiment 2a). All sentences were divided into five regions as exemplified in Table 7. These regions contained the locations of particular interest, namely the subject, verb and object region. Since the analysis of Experiment 1 revealed that the final 2-3 words of a sentence play a crucial role for the triggering of regressive eye movements, we divided the sentence part after the object into a spill-over and a final region. The final region contained the last two words unless they had a total length of less than 10 characters. In this case, we took the last three words.

As in Experiment 1, eye movement analysis focused on first pass time, go-past time and the probability of a first-pass regressive eye movement from a region. We again calculated linear mixed-effects models (for first pass time and go-past time) and logistic regression models (for probability of regressions out) for each region separately, using the maximal random effect structure justified by our design. In order to test for differences between the two experiments, we included experiment type (single vs. combined) as a fixed effect in our model, besides the effects of anomaly and association, as well as the interactions between these three factors. The models also included random intercept for subjects and items with random by-subject and by-item slopes for all the factors entered as fixed effects and their interaction (experiment type was only included in the by-item slopes). Again, we treat $|t|$ or $|z| > 2$ as significant.

Word-based eye-movement analysis

As in experiment 1, we carried out a second eye movement analysis that examined the pattern of regressive eye movements in more detail. We again restricted this word-based analysis to SRA sentences only and identified every single saccade, irrespective of whether it occurred in first pass or second pass reading or whether it went out of a particular region or not. Further, we excluded saccades that were triggered right before or after blinks was well as the last saccade in a sentence. Given our interest in inter-word saccades, we additionally excluded all saccades that landed within the same word (intra-saccades). In total, 34553 progressive and 9135 regressive saccades entered the word-based analysis.

6.2.4.2 Analysis of the neuroimaging data

The analysis of the neuroimaging data was by definition restricted to Experiment 2b. This analysis focused on the comparisons of the neural underpinnings of regressive vs. progressive eye movements by aligning the neuroimaging data with the eye-tracking data.

Analysis of the eye-tracking data

Because we were interested in comparing progressive and regressive saccades irrespective of sentence type, all inter-word saccades during sentence reading were identified and categorized. Saccades occurring right before or after blinks and saccade losses as well as the last saccade of a sentence were excluded from the analysis. In addition, we excluded all progressive saccades that were generated after a first regressive eye movement was performed, so that all progressive saccades undergoing analysis appeared in first pass reading.

There were generally more progressive than regressive saccades, ranging from a ratio of 2:1 to a ratio of 16:1. To keep the analysis of the two saccade types comparable and to increase the temporal spacing between individual progressive saccades, we selected progressive saccades to approximately match the number of regressive saccades at the single subject level. In total, 9897 progressive and 10691 regressive saccades were included in the analysis.

In addition, we computed the length of each saccade in characters, the on- and offset of each sentence and comprehension question, as well as trial-initial fixation periods. The analysis of eye tracking data was performed using python and bash scripts.

Analysis of the fMRI data

The analysis of the fMRI data was conducted using AFNI (https://afni.nimh.nih.gov/) and results were visualized using the surface mapping program SUMA (https://afni.nimh.nih.gov/Suma).

Pre-processing

The functional images underwent a standard pre-processing procedure (implemented in the afni_proc.py pipeline), consisting of de-spiking, slice-time correction, rigid-body motion correction, co-registration to the anatomical image and spatial normalization to the Talairach (TT_N27) template. The EPI time series were subsequently smoothed with a 6 mm FWHM Gaussian kernel and converted to percent signal change with a mean of 100. Time points that showed excessive motion (> 1 mm) were censored and subsequently excluded from regression analysis.

Single subject analysis

At the single subject level, we performed voxel-wise multiple linear regression using the program 3dREMLfit. The onsets of progressive and regressive saccades were included as separate regressors of interest and convolved with a single gamma function. Saccade length in characters was added as a continuous parametric regressor. The trial-initial fixation period and comprehension questions were collapsed into one regressor of no interest that was modelled with a duration modulated response model ('dmBLOCK'). The six motion parameters were included as nuisance regressors.

By including saccade length, we specified two response models per saccade type: The first model captures changes in BOLD activity correlated to the occurrence (i.e., onset) of saccades but not their variable length. The second model, however, indicates in which brain regions BOLD activity co-varies with changes in saccade length. For each subject, saccade length values were centered around the grand mean of progressive and regressive saccade lengths, respectively, to facilitate interpretation at the group level.

Group level analysis

The analysis at the group level was performed using the function 3dtest++ with the newly implemented function 'Clustsim' that derives cluster correction thresholds based on a nonparametric randomization and permutation algorithm. Importantly, this approach is independent of assumptions of the spatial autocorrelation distribution and has been shown to produce adequate global false positive rates (Cox, Chen, Glen, Reynolds, & Taylor, 2017).

One-sample t-tests were conducted for maps of mean activity per condition against the implicit baseline. Paired *t*-tests testing differences between progressive and regressive saccades were computed separately for the onset and saccade length model. Results were restricted to voxels inside a whole brain group mask and thresholded at a voxel-wise threshold of $p < .01$ and family-wise error rate of $p < .05$ to correct for multiple comparisons. For neuroanatomical labeling the brain atlas created by Destrieux and colleagues (Destrieux, Fischl, Dale, & Halgren, 2010) was used within SUMA.

6.2.5 Results

6.2.5.1 Question Accuracy

On average, 88.66% of all comprehension questions were answered correctly (SD: 5.36%, range 75.00 – 97.50%). There was no difference between experiments (mean Experiment 2a: 89.88%, mean Experiment 2b: 87.50%; $t(37.71)=1.44$, $p = .16$). These high accuracy results indicate that the subjects were reading carefully and for comprehension in both Experiments 2a and 2b.

6.2.5.2 Reading data

In the following, we will present the results of the two eye movement analyses for the SRA sentences, the region-based and the word-based analysis.

Region-based eye-movement analysis

The reading time measures as well as the statistical results of the region-based eye-movement analysis for the SRA sentences can be found in Table 8 and 9. The analysis revealed that the anomalous sentences led to increased go-past reading times and a higher probability for first pass regressions in all regions except the subject region. For first pass time, the anomaly effect was significant in the object and spill-over region but surprisingly not in the verb and final region. The lack of an anomaly effect in the verb region for first pass time especially stands in clear contrast to the results of Experiment 1, where a clear anomaly effect has been detected.

Association had a reliable impact on first past and go-past reading times in the object region, with longer reading times for low-associated compared to high-associated sentences. The same was true for go-past time in the final region where low-associated sentences also led to an increased probability for first-pass regressions.

In the object region, the anomaly and association effects in first pass time were modulated by a significant interaction between these two effects. Thus, the anomaly effect was reduced for high-associated compared to low-associated sentences. The same pattern was found for go-past time in the final region, again indicated by a significant interaction between anomaly and association.

Importantly, our analysis revealed no significant impact of experiment type on first pass reading or the probability to regress. Only for go-past time in the object region did an interaction of anomaly and experiment type become significant, showing that the anomaly effect was reduced for Experiment 2b compared to Experiment 2a.

Table 8: *By-subject means and standard deviations (in parentheses) of first pass time, go-past time and probability of regressions out for semantic reversal anomalies (SRA). Abbreviations: ano = anomalous sentences, non ano = non-anomalous sentences, single = single eye-tracking experiment, combined = combined fMRI / eye-tracking experiment, low = low associated, high = high associated.*

	Region 1 Subject *Der Lehrer / Der Fehler*				**Region 2 Verb** *korrigiert / errät*				**Region 3 Object** *den Fehler / den Lehrer*				**Region 4 Spill-Over** *nach zwanzig Berufsjahren schon*				**Region 5 Final** *fast automatisch.*			
	single		combined		single		combined		single		combined		single		combined		single		combined	
	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high
	First pass time																			
ANO	483 (129)	517 (149)	418 (88)	423 (121)	280 (100)	268 (92)	308 (126)	271 (100)	468 (110)	467 (140)	441 (120)	423 (155)	1169 (368)	1109 (329)	1178 (290)	1172 (333)	646 (194)	632 (216)	578 (166)	593 (195)
NON ANO	504 (137)	488 (133)	409 (114)	429 (120)	268 (114)	254 (89)	279 (107)	267 (81)	412 (101)	373 (93)	412 (106)	359 (105)	1224 (333)	1190 (260)	1221 (384)	1239 (392)	632 (174)	688 (167)	563 (170)	616 (174)
	Go-past time																			
ANO	529 (196)	556 (196)	452 (101)	473 (122)	350 (117)	319 (161)	373 (152)	353 (139)	636 (197)	598 (182)	566 (180)	549 (228)	1507 (388)	1435 (317)	1592 (486)	1546 (381)	1888 (779)	1561 (615)	1549 (858)	1281 (684)
NON ANO	532 (202)	533 (175)	467 (117)	482 (119)	298 (120)	286 (97)	331 (124)	327 (112)	469 (110)	422 (95)	468 (237)	430 (116)	1413 (347)	1322 (290)	1504 (466)	1399 (360)	1523 (605)	1345 (418)	1251 (638)	1191 (511)
	Regressions out																			
ANO					12.23 (9.98)	8.43 (10.24)	8.48 (7.40)	11.93 (11.37)	19.91 (15.92)	15.91 (11.85)	14.66 (12.35)	13.63 (11.29)	18.16 (14.43)	16.95 (12.88)	17.53 (12.97)	18.52 (15.35)	67.99 (21.16)	56.63 (22.04)	58.83 (22.86)	51.98 (21.47)
NON ANO					5.91 (6.13)	6.31 (7.66)	8.53 (10.21)	10.60 (12.85)	9.67 (11.84)	9.43 (7.99)	8.11 (5.95)	10.42 (9.13)	10.44 (8.96)	8.27 (6.28)	14.61 (8.02)	8.56 (10.84)	57.60 (28.83)	49.73 (18.99)	51.24 (25.17)	10.94 (26.07)

Table 9: *Results of the mixed models analysis (first pass time and go-past time) and the logistic regression models (regressions out) for semantic reversal anomalies (SRA).*

	Region 1 Subject *Der Lehrer / Der Fehler*			**Region 2 Verb** *korrigiert / errät*			**Region 3 Object** *den Fehler / den Lehrer*			**Region 4 Spill-Over** *nach zwanzig Berufsjahren schon*			**Region 5 Final** *fast automatisch*		
	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value	Estimate	Std. Error	|t| value / |z| value
First Pass Time															
Intercept	6.01	0.04	167.82	5.64	0.03	175.03	5.94	0.03	175.50	6.89	0.05	135.21	6.25	0.05	125.43
Anomaly	-0.01	0.03	0.19	0.03	0.02	1.56	0.11	0.03	**4.23**	-0.09	0.03	**3.06**	-0.05	0.03	1.79
Association	0.01	0.02	0.45	-0.04	0.02	1.76	-0.06	0.02	**3.14**	-0.03	0.04	0.69	0.02	0.03	0.63
Experiment	0.13	0.07	1.90	-0.08	0.06	1.28	0.00	0.06	0.01	0.00	0.08	0.05	0.06	0.08	0.73
Anomaly x Association	0.01	0.04	0.31	-0.02	0.03	0.58	0.10	0.04	**2.71**	-0.06	0.05	1.23	-0.06	0.06	0.95
Anomaly x Experiment	0.01	0.04	0.22	0.02	0.04	0.55	0.08	0.05	1.83	-0.04	0.05	0.76	-0.04	0.05	0.89
Association x Experiment	-0.03	0.04	0.66	0.03	0.04	0.64	0.03	0.04	0.65	-0.03	0.06	0.56	0.02	0.05	0.38
Anomaly x Association x Experiment	0.11	0.09	1.11	0.02	0.07	0.29	-0.07	0.08	0.87	0.07	0.11	0.65	-0.14	0.10	1.44
Go Past Time															
Intercept	6.12	0.04	146.06	5.78	0.04	163.43	6.11	0.04	165.84	7.15	0.05	141.78	7.00	0.06	114.65
Anomaly	-0.01	0.03	0.44	0.07	0.02	**2.73**	0.22	0.04	**6.16**	0.06	0.02	**3.82**	0.14	0.04	**3.23**
Association	0.02	0.02	1.10	-0.04	0.03	1.68	-0.06	0.02	**2.70**	-0.03	0.04	0.82	-0.14	0.04	**3.17**
Experiment	0.07	0.08	0.86	-0.13	0.07	1.95	0.00	0.07	0.06	-0.06	0.07	0.90	0.12	0.11	1.09
Anomaly x Association	0.01	0.04	0.18	-0.04	0.04	0.83	0.05	0.04	1.15	0.04	0.03	1.16	-0.15	0.07	**2.15**
Anomaly x Experiment	0.05	0.04	1.48	0.05	0.05	1.11	0.12	0.05	**2.16**	-0.02	0.03	0.62	0.06	0.08	0.74
Association x Experiment	0.00	0.03	0.00	-0.03	0.05	0.62	-0.03	0.04	0.77	-0.01	0.03	0.19	-0.01	0.06	0.11
Anomaly x Association x Experiment	0.01	0.08	0.13	-0.04	0.08	0.52	-0.10	0.08	1.16	-0.03	0.07	0.49	-0.07	0.10	0.68
Regressions Out															
Intercept				-2.98	0.17	17.86	-2.27	0.13	17.34	-2.27	0.14	16.05	0.27	0.16	1.67
Anomaly				0.47	0.18	**2.54**	0.52	0.14	**3.70**	0.62	0.18	**3.39**	0.37	0.12	**3.00**
Association				0.13	0.17	0.77	0.17	0.15	1.11	-0.17	0.15	1.12	-0.35	0.12	**2.98**
Experiment				-0.44	0.30	1.45	0.09	0.26	0.33	-0.19	0.27	0.68	0.26	0.32	0.81
Anomaly x Association				-0.17	0.43	0.38	-0.04	0.27	0.14	0.49	0.26	1.88	-0.34	0.19	1.78
Anomaly x Experiment				0.64	0.34	1.91	0.29	0.27	1.06	0.22	0.34	0.66	0.21	0.22	0.95
Association x Experiment				-0.34	0.33	1.06	-0.29	0.28	1.02	0.30	0.28	1.08	-0.32	0.20	1.56
Anomaly x Association x Experiment				-0.48	0.84	0.57	0.08	0.54	0.16	-0.66	0.52	1.26	0.07	0.39	0.19

Word-based eye-movement analysis

As for Experiment 1, our word-based eye-movement analysis of the SRA sentences focused upon the landing site pattern of regressive eye movements and the fixation durations, before a regressive or progressive saccade respectively was initiated.

Distribution of regression landing sites

We started inspecting the regression landing site pattern as in Experiment 1 by computing the length of all regressive saccades. Figure 13 shows the distribution of regression length for regressions that were initiated in the final region of the sentence.

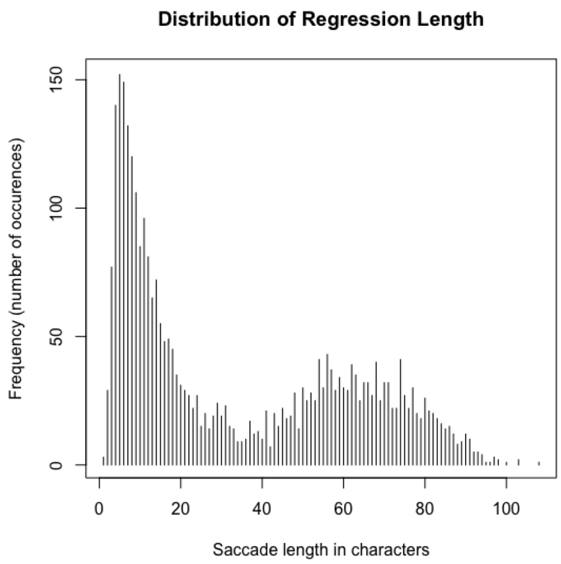

Figure 13: *Distribution of Regression Length for the German SRA sentences. Shows the number of occurrences (y-axis) of all regressions that were initiated in the final region (n=3190) with the corresponding length in characters (x-axis)*

These regressions had a mean length of 33.64 characters (SD: 27.76). The visual inspection shows that again the majority of regressions was short, but the proportion of regressions with a length of 15 characters or less dropped to 24.46%. However, for all regressions the proportion of saccades with a length of 15 characters or less remained high (68.34%). In comparison to Experiment 1, a second peak can be identified, so that also a great proportion of regressions from the sentence final region had a length of 55–75 characters (19.00 %). To examine this pattern in more detail, we computed the target position of regressions from the sentence beginning (see Figure 14), again only for regressions that were initiated in the final region. Here the two peaks became very apparent: one group of regressions targeted the sentence beginning, whereas a second group targeted a position much closer to the current fixation. This pattern is a clear replication of the findings of Experiment 1: Regressions do not all behave in the same way but can be further assigned to two groups depending on their target position.

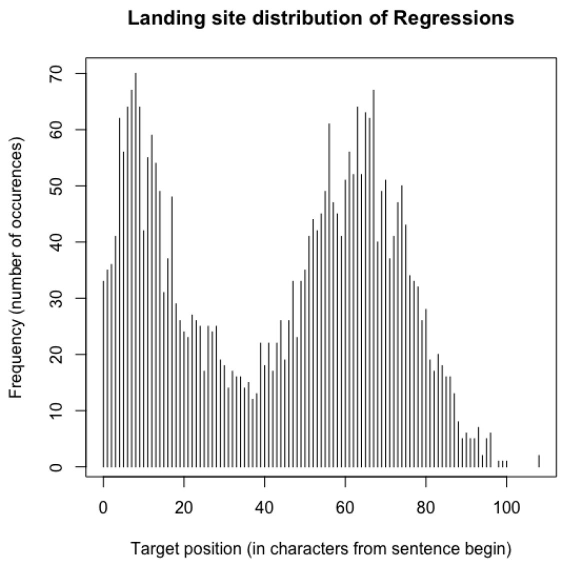

Figure 14: *Landing site distribution of Regressions. Shows the number of occurrences of regressions that were initiated in the final region for a particular position from the sentence beginning (in characters).*

Since the target position in characters is not very informative with regard to the sentence region targeted by the regressions (because the sentences varied with respect to their word and sentence length), we aligned the target positions with the sentence regions. For this analysis, we used the regioning scheme that is exemplified in Table 7 but made eight regions instead of five (see Table 10) to get a more fine-grained picture of the results. To achieve this, we further divided the subject and object region into the determiner and the noun region. This was done because the determiner in German carries the case marking (in contrast to English) and therefore provides important information for a potential reanalysis of the sentence. In addition, we divided the original spill-over region into a very short new spill-over region, consisting of only one word (unless it was longer than 2 characters, in the other cases we took two words) and the remaining pre-final region. The verb and final region remained the same.

Table 10: *Alternate regioning-scheme for the German SRA sentences.*

Region 1 Subject - determiner	**Region 2** Subject - noun	**Region 3** Verb	**Region 4** Object - determiner	**Region 5** Object - noun	**Region 6** Spill-over	**Region 7** Pre-final	**Region 8** Final
Der	*Lehrer*	*korrigiert*	*den*	*Fehler*	*nach*	*zwanzig Berufsjahren schon*	*fast automa-tisch.*

We computed the proportion of regressions of all inter-region regressions that targeted a particular region. In addition, we were interested whether the target positions differed between sentence types (anomalous vs. non-anomalous) as is expected by the *Information Gathering Framework* and the results of Experiment 1. Thus, we computed two-sample chi-squared tests between the proportions of these two sentence types for each sentence region[34].

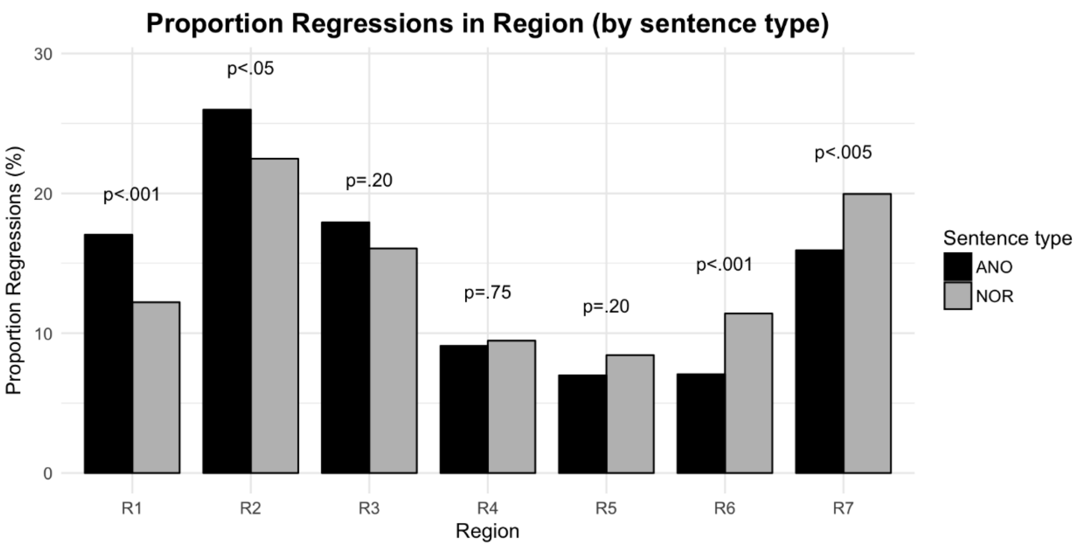

Figure 15: *Proportion of Regressions in Region by sentence type. Shows the distribution of all inter-region regressions (proportion from all inter-region regressions of a sentence type in percent) of a particular sentence type with regard to the region where the regression landed. Chi squared tests compared the difference of proportions between sentence types for each region. Abbreviations: ANO = anomalous sentences, NOR = non-anomalous sentences, R1 = subject-determiner, R2 = subject-noun, R3 = verb region, R4 = object-determiner, R5 = object-noun, R6 = spill-over, R7 = pre-final.*

[34] Note again that it is problematic to use chi-squared tests to compare the proportions of regressions for each region separately. However, due to the lack of alternatives, we used this method but we are also aware that the results of the interferential statistics have to be interpreted with caution.

Figure 15 shows the distribution of the target positions for the regressions by sentence type. The analysis revealed that regressions in anomalous sentences targeted significantly more often in the subject region, and the effect size shows that the difference was even greater for the determiner ($\chi^2 = 13.61$, *df* = 1, p<.001, *Cramér's V* = 0.15) than for the noun ($\chi^2 = 4.00$, *df* = 1, p<.05, *Cramér's V* = 0.06). For the verb and the object-region no significant differences were found. In the spill-over and pre-final region, the proportion of regressions into this region was smaller for anomalous sentences. These results indicate that the target position of regressions is highly affected by higher linguistic constraints such as syntax and semantic processing.

We further examined whether experiment type also affected the landing site patterns and carried out the same analysis for regressions from Experiment 2a vs. regressions from Experiment 2b. However, our analysis revealed no differences for the subject, verb, spill-over or pre-final region (all $\chi^2 < 1.60$). Only in the object region was there a small effect of experiment type, so that regressions in the single experiment were more likely to target the object-determiner ($\chi^2 = 4.96$, *df* = 1, p<.05) and less likely to target the object-noun ($\chi^2 = 4.86$, *df* = 1, p<.05).

Fixation duration before regressions

As in experiment 1, we also examined the fixation durations before regressive vs. progressive eye movements. Recall that the *Information Gathering Framework* makes very clear predictions with regard to fixation durations, namely, that fixations that precede regressive eye movements should be shorter on average but that this should not be the case in the sentence final region.

In order to test this hypothesis, we carried out linear mixed-effects models for log fixation durations with saccade type, anomaly and experiment type as well as their interactions as fixed effects for each of the sentence regions separately. We used the maximal random-effects structure for the by-subjects and by-item slopes (experiment type was only included into the by-item slopes). Table 11 provides the results of the statistical analysis and Figures 16 and 17 visualize the fixation durations for each experiment type separately.

Table 11: *Results of the linear mixed-effects models for log fixation duration by sentence region. Note that there were no regressions initiated in first region because here the first region consisted of only one word.*

	Region 2 Subject – noun			**Region 3** Verb			**Region 4** Object – determiner			**Region 5** Object – noun			**Region 6** Spill-over			**Region 7** Pre-final			**Region 8** Final		
	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value	Estimate	Std. Error	|t| value
Intercept	5.37	0.02	298.81	5.39	0.02	314.93	5.41	0.02	307.19	5.39	0.02	349.20	5.40	0.02	287.79	5.39	0.01	391.00	5.51	0.02	285.36
Anomaly	0.00	0.02	0.16	0.05	0.02	**2.78**	0.06	0.03	**2.01**	0.04	0.02	**2.10**	0.02	0.03	0.53	0.01	0.01	0.40	-0.02	0.01	1.06
Saccade type	0.19	0.03	**7.07**	0.21	0.02	**9.11**	0.20	0.02	**8.11**	0.15	0.02	**7.20**	0.18	0.03	**5.67**	0.12	0.02	**8.10**	0.01	0.03	0.56
Experiment type	0.14	0.04	**3.96**	0.10	0.03	**2.97**	0.16	0.03	**4.74**	0.14	0.03	**4.40**	0.11	0.04	**2.63**	0.15	0.03	**5.50**	0.11	0.04	**3.00**
Anomaly x Saccade type	0.02	0.04	0.50	-0.03	0.03	0.79	-0.02	0.06	0.31	-0.04	0.03	1.10	0.04	0.06	0.66	0.02	0.03	0.70	0.04	0.03	1.72
Anomaly x Experiment type	0.02	0.04	0.60	-0.01	0.03	0.20	-0.06	0.05	1.26	0.04	0.03	1.30	0.05	0.05	0.92	-0.01	0.03	0.50	0.01	0.03	0.53
Saccade type x Experiment type	0.06	0.06	1.02	0.10	0.05	**2.06**	0.06	0.05	1.32	0.15	0.04	**3.60**	0.08	0.07	1.26	0.08	0.03	**2.60**	0.15	0.05	**3.01**
Anomaly x Saccade type x Experiment type	-0.05	0.08	0.62	0.02	0.06	0.33	-0.06	0.10	0.63	-0.11	0.07	1.60	-0.07	0.11	0.61	0.03	0.05	0.60	0.01	0.05	0.29

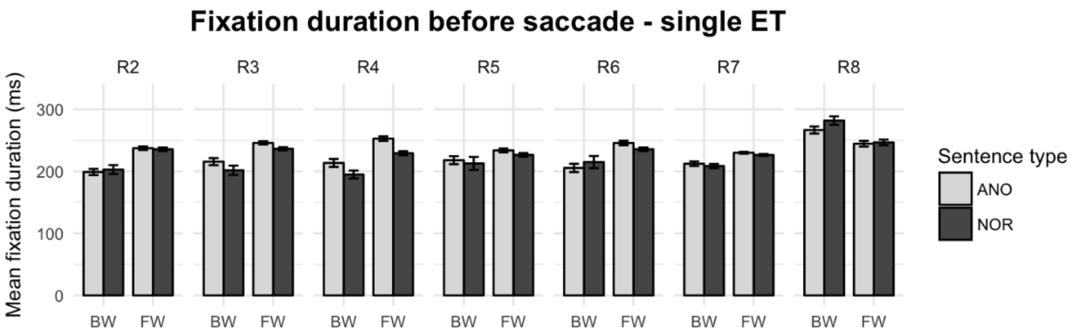

Figure 16: *Fixation duration before saccade for Experiment 2a. Mean fixation durations and standard errors are given for each saccade type and sentence type separately. Abbreviations: BW = regressive saccade, FW = progressive saccade, ANO = anomalous sentences, NOR = non-anomalous sentences, R2 = subject-noun, R3 = verb, R4 = object-determiner, R5 = object-noun, R6 = spill-over, R7 = pre-final, R8 = final region. See text and Table 11 for further information.*

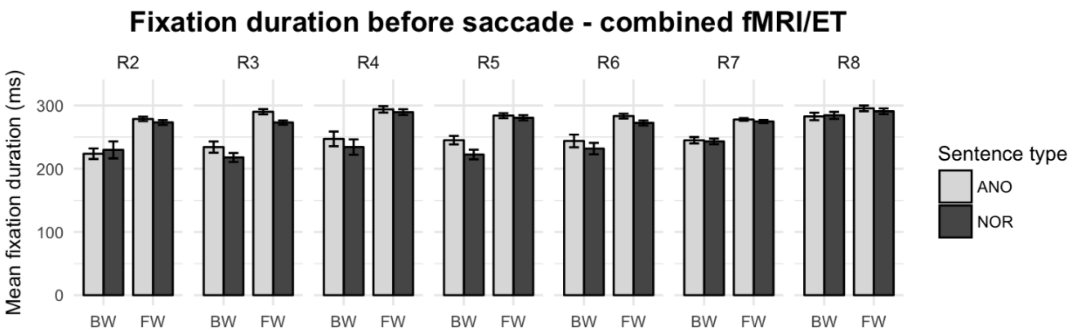

Figure 17: *Fixation duration before saccade for Experiment 2b. Mean fixation durations and standard errors are given for each saccade type and sentence type separately. Abbreviations: BW = regressive saccade, FW = progressive saccade, ANO = anomalous sentences, NOR = non-anomalous sentences, R2 = subject - noun, R3 = verb, R4 = object - determiner, R5 = object - noun, R6 = spill-over, R7 = pre-final, R8 = final region. See text and Table 11 for further information.*

As predicted by the *Information Gathering Framework,* we found a reliable effect of saccade type for all regions except the last region which counts for the reduced fixation durations before a regressive saccade was initiated. The general pattern and the lack of the saccade type effect in the last region replicates the findings of Experiment 1 and is in accordance with the predictions of the *Information Gathering Framework*.

In addition to the effect of saccade type, we found a reliable impact of experiment type for all regions, meaning that fixations were in general longer in the combined fMRI / eye-tracking experiment. This is a surprising result because we found no influences of experiment type in the region-based eye-movement analysis (despite the interaction with anomaly in go-past time for the object region).

Also in contrast to the results from Experiment 1, the analysis revealed a significant effect of anomaly on fixation duration for the verb and object region with longer fixation durations for anomalous sentences.

Finally, we found a significant interaction of saccade type and experiment type for the verb, object-noun, pre-final and final region. This interaction accounts for the fact that the difference between fixation durations before progressive and regressive saccades was smaller for the single eye-tracking experiment for the verb, object-noun and pre-final region, but this pattern was reversed in the final region.

6.2.5.3 Functional Imaging

Comparisons against baseline

We first analyzed changes in BOLD signal correlated to the onset and length of each of the two saccade types relative to implicit baseline. A detailed description of the resultant clusters is provided in Table 12 and shown in Figure 18.

Table 12: *Activation cluster in comparison to implicit baseline for regressive and progressive saccades. The table contains an anatomical description, the Talairach coordinates of the local maximum and the voxel size of the significant clusters for saccade onset (threshold: 125 voxel) and the parametric regressor of saccade length (threshold 113 voxel) for each saccade type separately.*

Hemisphere	**Cluster extent**	**x**	**y**	**z**	**voxel**
Regressive saccades (onsets) > Baseline					
Positive correlation					
Right	Superior frontal gyrus and sulcus, spreading into the middle-anterior part of the cingulate gyrus and sulcus, the middle frontal gyrus and sulcus, the opercular part of the inferior frontal gyrus and the inferior frontal sulcus. Extending parietal to precentral gyrus, intraparietal sulcus and transverse parietal sulci, superior and middle occipital gyri, superior occipital sulcus and transverse occipital sulcus, including also parts of the precuneus, subparietal sulcus and parieto-occipital sulcus.	40	-17	7	2590
Left	Superior frontal gyrus and sulcus, spreading into anterior and the middle-anterior part of the cingulate gyrus and sulcus, the middle frontal gyrus and sulcus, the inferior and superior part of the precentral sulcus and inferior frontal sulcus. Triangular and orbital part of the inferior frontal gyrus, short insular and orbital gyri, the supramarginal gyrus, the intraparietal sulcus and transverse parietal sulci as well as the precuneus.				
Right	Lingual gyrus and sulcus, lateral occipito-temporal gyrus	-33	57	-25	329
Left	Lingual gyrus and occipital pole				

Right	Cerebellum	23	50	39	254
Left	Cerebellum				
Negative correlation					
Right	Suborbital sulcus, anterior parts of the superior frontal gyrus and the cingulate gyrus and sulcus	-2	-55	7	325
Left	Transverse frontopolar gyri and sulci, anterior parts of the superior frontal gyrus and the cingulate gyrus and sulcus, subcallosal area				
Progressive saccades (onsets) > baseline					
Positive correlation					
Right	Precentral gyrus, spreading into central sulcus, postcentral gyrus and the superior part of the precentral gyrus	-37	29	59	206
Negative correlation					
Right	Anterior parts of the superior frontal gyrus, anterior and middle-anterior part of the cingulate gyrus and sulcus, middle frontal sulcus and gyrus, transverse frontopolar gyri and sulci, pericallosal sulcus	-19	-55	24	317
Left	Anterior and middle-anterior part of the cingulate gyrus and sulcus				
Regressive saccades (length) > Baseline					
Positive correlation					
Right	Lingual gyrus, occipital pole, superior occipital gyrus and sulcus and transverse occipital sulcus	16	71	8	250
Left	Lingual gyrus, occipital pole and middle occipital pole				
Left	Middle frontal sulcus and gyrus, fronto-marginal gyrus and sulcus	30	-38	21	218
Right	Subparietal sulcus and precuneus	9	57	7	724
Progressive saccades (length) > Baseline					
Positive correlation					
Right	Superior occipital gyrus and sulcus an transverse occipital sulcus, cuneus and precuneus, parieto-occipital sulcus, lingual gyrus, lateral occipito-temporal gyrus and posterior transverse collateral sulcus	-19	57	7	724
Left	Superior occipital gyrus, superior occipital sulcus and transverse occipital sulcus, middle occipital sulcus and lunatus sulcus, occipital pole, cueneus, calcarine sulcus and lingual gyrus				
Right	Middle-anterior and middle-posterior part of the cingulate gyrus and sulcus, superior frontal gyrus	-12	20	35	122
Left	Middle-anterior and middle-posterior part of the cingulate gyrus and sulcus as well as the middle-anterior superior frontal gyrus				

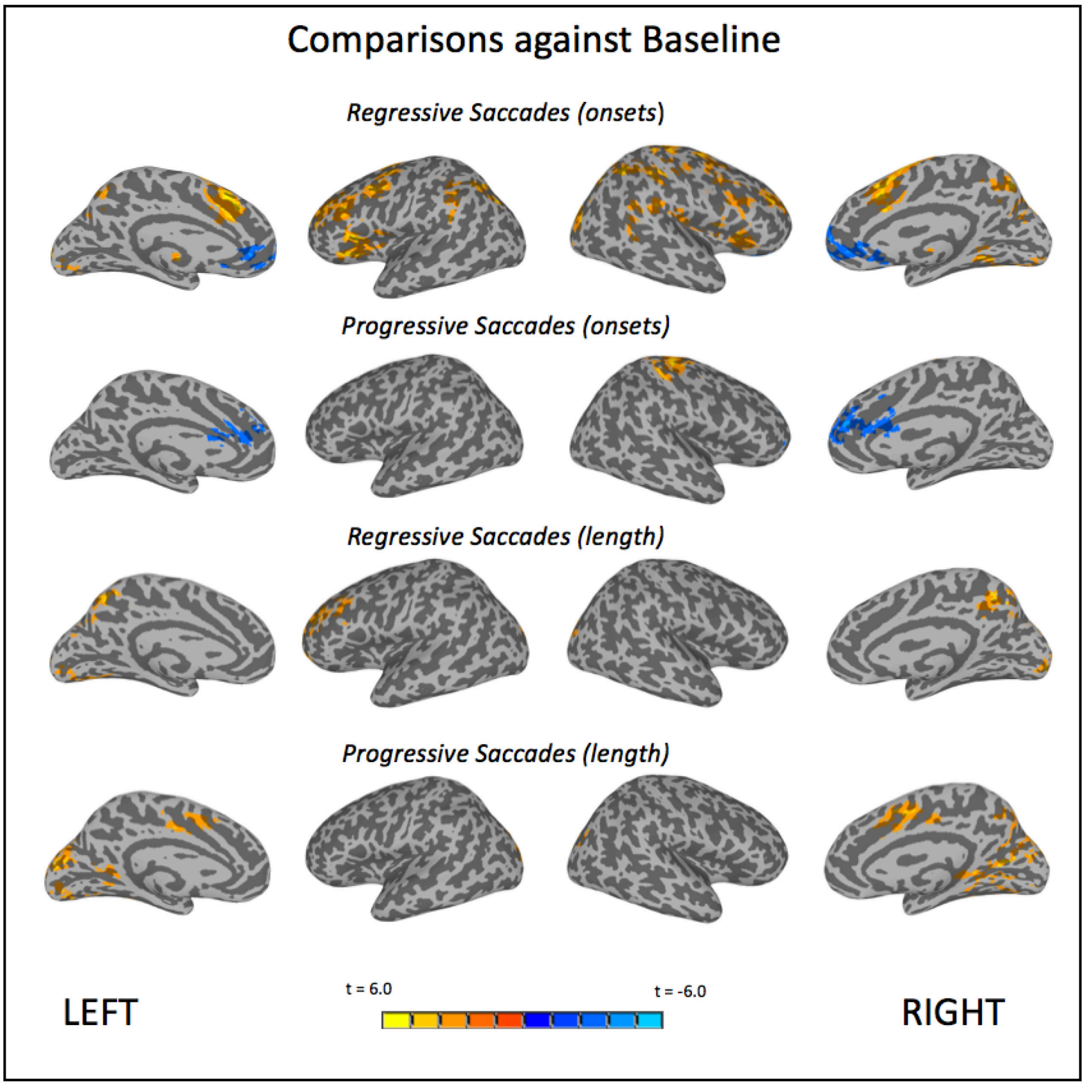

Figure 18: Areas of activation significantly correlated with the onset and saccade length, respectively, of regressive and progressive saccades in comparison with implicit baseline.

For the comparison of the onset of regressive eye movements with the baseline, the analysis revealed large clusters of increased activation across both hemispheres. Increased levels of BOLD activity were found bilaterally in the middle frontal gyrus (MFG) and sulcus (MFS), the medial superior frontal gyrus (SFG) and sulcus, as well as the middle-anterior part of the cingulate gyrus and sulcus (aMCC) and the right precentral gyrus. Occipital areas also showed increased levels of activity in the right superior occipital gyrus (SOG) and the left occipital pole as well as in the left cerebellum. Bilateral decreases in BOLD activity were found in frontal areas including the anterior parts of the cingulate gyrus and sulcus (ACC), as well as in the right anterior SFG.

For the onset of progressive eye movements, we observed an increased involvement of the right precentral gyrus and decreased involvement of the right and left ACC, as well as the anterior parts of the right SFG compared to implicit baseline.

For regressive saccades, the analysis of changes in BOLD signal correlated to saccade length revealed increased levels of BOLD activity for longer saccades in the left MFG and MFS, as well as in occipital regions including the right SOG and the bilateral occipital pole. For progressive saccades, saccade length was also positively correlated with bilateral activations in the SOG and the cuneus, as well as in the aMCC.

Regressive versus progressive saccades

Our main analysis, however, focused on the direct comparison of neural changes related to the onset and length of regressive versus progressive saccades. The results are presented in Table 13 and Figure 19.

Table 13: *Activation cluster for comparison of regressive and progressive saccades. The table contains an anatomical description, the Talairach coordinates of the local maximum and the voxel size of the significant clusters for saccade onset (threshold: 125 voxel) and the parametric regressor of saccade length (threshold 113 voxel) for each saccade type separately.*

Hemisphere	**Cluster extent**	**x**	**y**	**z**	**voxel**
Regressive saccades > Progressive saccades (onsets)					
Positive correlation					
Right / Left	Medial superior frontal gyrus, extending inferior to the middle-anterior part of the cingulate gyrus and sulcus	-5	-17	35	360
Right	Middle frontal sulcus, spreading to transverse frontopolar gyri and sulci, middle frontal gyrus, inferior frontal sulcus and the inferior part of the precentral sulcus	-33	1	35	329
Right	Intraparietal sulcus and transverse parietal sulci as well as the postcentral sulcus, spreading to the sulcus intermedius primus (of Jensen), superior temporal sulcus as well as parts of the supramarginal gyrus	-44	29	42	297
Left	Anterior middle frontal gyrus and sulcus, extending superior into the anterior superior frontal sulcus and inferior to the medial inferior frontal sulcus, including the triangular part of the inferior frontal gyrus	30	-24	31	189
Left	Intraparietal sulcus and transverse parietal sulci, including medial parts of the postcentral sulcus and gyrus; supramarginal gyrus with extensions to the sulcus intermedius primus (of Jensen) and the angular gyrus.	33	50	38	185

Progressive saccades > Progressive saccades (length)					
Negative correlation					
Right	Superior occipital gyrus, extending to the superior occipital sulcus and transverse occipital sulcus, the cuneus, the parieto-occipital sulcus and the precuneus	2	82	14	154
Left	Cuneus, extending to the superior occipital gyrus, the middle occipital sulcus and the lunatus sulcus as well as the calcarine sulcus				

For the contrast of the BOLD changes correlated with the onset of regressive versus progressive saccades, we found increased levels of BOLD activity for regressions compared to progressions in the right MFG and MFS, the opercular part of the IFG, the medial SFG, the ACC and aMCC as well as in the intraparietal sulcus (IPS), and the anterior intermediate parietal sulcus ('sulcus of Jensen'). In the left hemisphere, regression onsets revealed higher levels of BOLD activity in the medial SFG, the ACC and aMCC, the MFG and MFS, as well as in the IPS and the supramarginal gyrus (SMG).

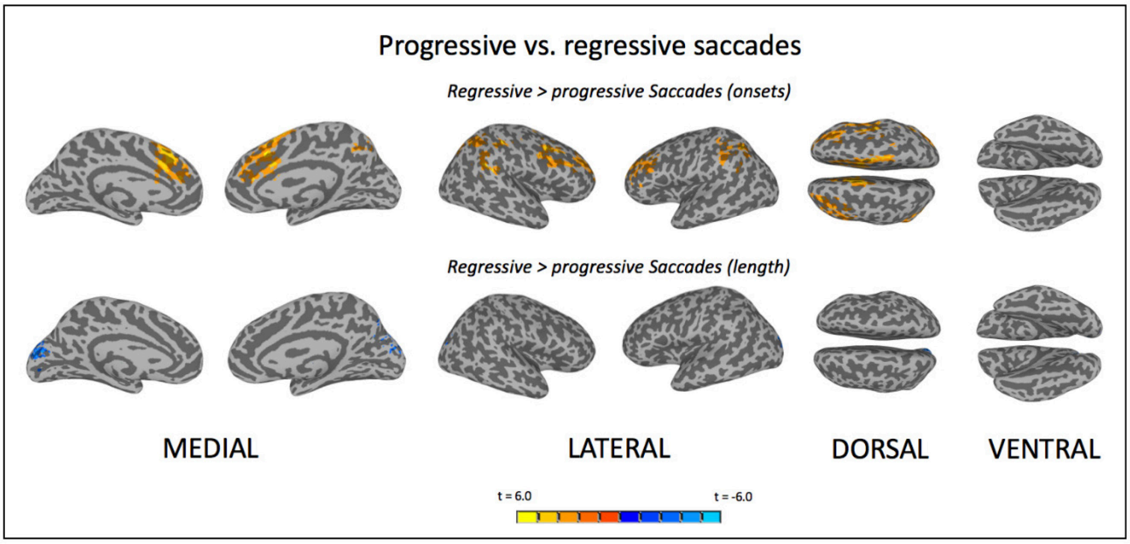

Figure 19: *Areas of activation significantly correlated with the onset and saccade length, respectively, for the comparison of regressive and progressive saccades.*

The contrast on co-modulation of neural activity with saccade length revealed decreased levels of BOLD activity for regressive compared to progressive eye movements in occipital areas like the cuneus and the SOG, bilaterally.

6.3 Discussion

In the current experiments, we examined the neural underpinnings of regressive and progressive eye movements during free reading of different sentence types in a combined fMRI / eye-tracking experiment. In addition, we compared the eye movement patterns that were recorded in the scanner with those recorded outside the scanner in order to test the impact of the two experimental settings. This gave us additional insights into the role of regressive eye movements during reading of anomalous and non-anomalous sentence material such as German SRA sentences.

The comparison of the neural correlates of regressive saccades with progressive saccades revealed qualitatively different activation patterns for these two types of saccades. Whereas progressive saccades showed deactivations in the anterior prefrontal cortex compared to baseline, regressive saccades elicited a great activation network in both hemispheres including the medial SFG, the ACC, IPS and DLPFC. We also found that the length of a saccade correlated with increased activation in visual areas for both types of saccades and in the anterior prefrontal cortex for regressive saccades in particular.

The eye movement results showed that the type of experiment (single [2a] vs. combined [2b]) had no impact on first pass reading time or regression probability for SRA sentences. In addition, we found a reliable effect of anomaly in the verb and object region in all measures, except first pass time in the verb region. Also, a high association between the verb and the inanimate noun led to shorter reading times on the object region for both first and go-past time. The examination of fixation durations bear out the fact that despite a general finding of reduced fixation duration before regressions, fixations were also longer for Experiment 2b compared to Experiment 2a overall.

In the following we will discuss these findings in turn, starting with the eye movement results.

6.3.1 Eye-tracking results

Because Experiment 1 showed that a manipulation of comprehension question difficulty had a strong impact on the regression behavior at the end of a sentence, we asked whether a different experimental setting would also affect reading behavior. We therefore compared the reading behavior of subjects lying in a fMRI scanner with those sitting more or less comfortably on a chair and reading sentences on a computer screen. As for the question difficulty manipulation of Experiment 1, we did not find any impact of experiment type on first pass reading behavior. In addition and in clear contrast to Experiment 1, however, experiment type did not affect the probability of regressions, either. Only the anomaly effect was reduced for the combined fMRI / eye-tracking experiment in the object region (although it was still present), which indicates that people were spending less time rereading the object region. We will discuss this particular pattern below in the context of the single fixation durations.

Surprisingly, our word-based analysis revealed that experiment type nevertheless had a general impact on reading behavior although it was not visible in the region-based analysis. Specifically, single fixation durations were always longer for the combined fMRI / eye-tracking experiment than for the single eye-tracking experiment. This of course raises the question as to why increased single fixation durations did not lead to increased first pass reading times. Obviously, the subjects made more but shorter fixations in Experiment 2a, which led in total to approximately the same amount of time spent in a region as in Experiment 2b.

One explanation for this could be that the eye-tracker in the scanner was not as accurate as the one outside the scanner (i.e., combining several shorter fixations into one longer fixation), although both were from the same company (SR Research) and had the same high resolution of 1000 Hz. But in contrast to the tower-mount system used in the single eye-tracking experiment, the eye-tracker in the scanner had to correct for (small) head movements as well. Most importantly, however, the distance from the eyes to the monitor was greater in the scanner, perhaps leading to less precise measurements.

However, it is also possible that the subjects in the scanner in fact applied a different basic reading behavior with less single fixations, probably due to the more unfamiliar reading environment. In any case, since this reading behavior did not affect first pass reading times nor sentence interpretation, we argue in terms of our *Information Gathering Framework* that experiment type did not adjust the forward threshold. We rather think that the results emphasize the fact that fixations during sentence reading can be best viewed at least as word-driven, but maybe better as even region-driven: Thus, the shape of the eye movement pattern during sentence and text reading seems to be best explained by word-based, or even region-based accounts rather than single fixation accounts.

In sum, the results show that experiment type did not affect sentence reading behavior on first pass, which implies that it adjusted neither the forward nor the backward threshold in terms of our *Information Gathering Framework,* in contrast to the question difficulty manipulation in Experiment 1. This is a very important finding because it emphasizes the reliability of the eye-tracking data acquired in the scanner and has methodological implications for future research. Thus, the results indicate that eye-movement data from combined fMRI / eye-tracking experiments can in fact be used to explore the neural correlates of natural reading. In addition, it shows for a further time that the basic mechanisms of sentence interpretation during reading are largely unaffected by experimental settings like task or environment.

Another focus in our analysis of the eye-tracking data was set upon the examination of fixation durations before regressions. Our results clearly replicate the findings from Experiment 1 as well as from the literature, and further support the assumptions of the *Information Gathering Framework*. Again, we were able to show that fixations preceding regressive eye movements were of shorter duration than those preceding progressive eye movements. In particular, whereas this was true for all regions, we once again found evidence that fixations occurring before regressions at the end of a sentence (which we view as a response to missing expected evidence) were as long as their progressive counterparts. This is a finding which to our knowledge has never been previously reported in the literature and which can be well explained by the *Information Gathering Framework*.

In contrast to Experiment 1, we also found (despite the significant effect of experiment type, see above) that anomaly led in general to increased fixation durations for anomalous sentences in Region 3 (verb), 4 (object–determiner) and 5 (object–noun), although the effect was not very strong compared to the effects of saccade and experiment type. This finding is nonetheless not unexpected, given the fact that reading times for these regions were also increased. However, one could argue that in terms of the *Information Gathering Framework,* fixations in anomalous sentences should be shorter than in non-anomalous sentences because they reflect integration difficulties. Note however, that *the Information Gathering Framework* makes the prediction that only fixations before regressions should be shorter compared to progressions when these regressions are caused by integration difficulties. And this is the exact pattern we found in our data, because the effect of sentence type (anomalous vs. non-anomalous) did not override the effect of saccade type. In addition, the forward threshold (which triggers a regression in the case of integration difficulties) is not defined as a time threshold but as a threshold of confidence level. Thus, for anomalous sentences it often (but not necessarily) happens that the computation of the confidence level already reveals some problems in the lower linguistic levels, which leads to increased fixation durations. Since we did not find any reliable effects of sentence type on fixation durations in Experiment 1, this could suggest that the processing difficulties on the single fixation level were simply not strong enough to increase single fixation durations in general although these effects were clearly visible in first pass time.

In addition to the general impact of anomaly, we found that the anomaly effect was reduced for the combined fMRI / eye-tracking experiment, indicated by a significant interaction with experiment type in region 3 (verb), 5 (object–noun), 7 (pre-final) and 8 (final). In the object region, this interaction was also found in go-past time in the region-based analysis. Together with the lack of an interaction in first pass regression probability in all regions, this suggests that these reduced fixation durations are probably due to a shorter rereading of these regions instead of a general reduced probability to reread (i.e., by making less regressions). Thus, subjects in the combined fMRI / eye-tracking experiment appeared to spend less time in rereading sentence material when faced with anomalous sentences compared to subjects in the single eye-tracking experiment. It could be that this was just because the subjects in the scanner wanted to finish the experiment earlier. Note, however, that the effect was still small and failed to reach significance in the region-based analysis for region 4 (spill-over) and 5 (final). Thus, we should consider these results with caution. Importantly, experiment type did not influence the detection of an anomaly because the anomaly effect was still present in all regions. Consequently, they do not question the general reliability of the eye-tracking data, because language processing was not affected (note also – although we did not examine this in more detail here – there were no differences with regard to comprehension accuracy between the two experiment types). Nor do the results provide any difficulties for the *Information Gathering Framework* because the model is not specified with regard to second pass reading as yet.

In our experiments we also employed German counterparts of the English SRA sentences used in Experiment 1 to compare the eye movement pattern and language processing mechanisms between these two languages. Our data replicates the finding from Experiment 1: Even a high association between verb and object did not eliminate the detection of the anomaly. Thus, the anomaly effect was clearly visible in the object region, although a sentence structure with a reversal of thematic roles (likely to induce a plausible interpretation of the anomalous sentences if the reader adopted a superficial reading strategy) was chosen. These results again suggest that sentence interpretation is always complex and not underspecified on the syntactic level, which we interpret as a hint against the assumptions of the good enough approach that proposes that readers adopt a sentence interpretation that is just "good enough" for the task at hand. But note that in contrast to Experiment 1, we did not explicitly manipulate the task (unless we view the experiment type as some kind of task). Thus, the predictions of the good enough approach for Experiment 2 are not as strong as for Experiment 1.

Besides these replications, our experiments also revealed some important differences between the processing of the English and German SRA sentences. In particular, whereas the reading times on the verb region were increased for the English SRA sentences, we were not able to find any such effect for their German counterparts in first pass time. An initial interpretation might be that in the German sentences the anomaly effect was delayed and not detected in first pass time on the verb. However, this interpretation can hardly be correct as our analysis also showed an increased probability for first pass regressions out of the verb region. It is therefore possible to conclude that the detection of the anomaly is not delayed but rather occurs earlier in German than in English. Thus, the early detection of the anomaly may have caused an early regressive eye movement without increasing fixation duration. This dissociation between increased fixation duration and regression behavior is one of the core principles of the *Information Gathering Framework* and fits well to the findings of reduced fixation durations before regressions which has been discussed above.

However, if this assumption holds true, we have to ask the question as to why there is an earlier detection of the anomaly in German than in English. We consider a plausible explanation may be that the two languages differ with regard to their underlying form-to-function mappings (MacWhinney et al., 1984). Thus, in English, the actor of a sentence can primarily be identified on the basis of word order information, whereas in German case marking and animacy are more reliable to identify the actor. If we apply these considerations to the SRA sentences used in our studies, we should expect different processing patterns, especially on the verb.

Specifically, in both languages the inanimate initial argument (*the flower / der Fehler*), which is unambiguously marked as the actor (either by word order as in English or by case marking as in German), causes an animacy violation on the verb position because the verb requires an animate actor. In German, however, actor identification relies primarily on case marking and animacy. Thus, these information cues are evaluated quite early in the argument interpretation and consequently lead to an early error signal. In English by contrast, the animacy information is not used to develop a first interpretation of the argument because sentence interpretation primarily relies on word order information and this information does not lead to any difficulties. In only a later step, the sentence interpretation is matched with world knowledge and an error signal is produced, resulting in longer reading times or regressions. These results are therefore in line with crosslinguistically motivated language processing models such as the extended Argument Dependency Model (eADM; Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009) which claim a different time course in argument interpretation for English and German (see also Bornkessel-Schlesewsky & Schlesewsky, 2008, for a discussion of crosslinguistic differences in ERP effects elicited by SRA sentences).

More evidence for different processing patterns between English and German is gleaned from go-past time on the object position. Whereas in both languages we found an almost identical reading pattern in first pass time with only slightly reduced reading times for high compared to low associated anomalous sentences, this pattern diverged in go-past time. In English, the reduced first pass reading times actually led to increased go-past times which suggest that the anomaly effect was detected later for high associated sentences compared to low associated sentences. In German by contrast, a high association generally led to reduced reading times and the size of the anomaly effect was not shaped. This indicates that in German the anomaly effect was detected regardless of lexical association but based on animacy evaluations. In English by contrast, a high association led to a delayed effect of anomaly due to the important role of word order.

Finally, our eye-tracking analysis was able to replicate findings from Experiment 1 with regard to the regression target pattern. Again, we found evidence that the majority of regressive eye movements were short, although the results were not as clear as in Experiment 1. Nevertheless, two different target patterns of regressions in the sentence were again apparent. One group of regressions targeted the sentence beginning whereas the second group targeted a position much closer to the current fixation position. This suggests that regressions in general may aim to either reread the sentence from the beginning (c.f. the rereading pattern from von der Malsburg & Vasishth, 2011, 2013; and the forward reanalysis strategy proposed by Frazier & Rayner, 1982) or to reread a portion that is within 15-20 characters to the left of the current fixation. Although still more research is needed to examine the landing site distributions of regressions in more detail, these results are generally in line with the *Information Gathering Framework* that differentiates between a strategy based target selection and a language based target selection whereas the latter can only be applied for words within the perceptual span (which comprises about 15 characters).

In addition, the inspection of the landing site pattern again revealed evidence for a linguistically controlled target selection. Thus, we found increased re-inspections of the subject and verb region in anomalous sentences. Given that these regions cause the integration difficulties because they constitute the animacy violation, this is a very reasonable regression pattern. Note, however, that the subject region was also the initial region of the sentence. Thus, it is not possible to distinguish regressions that aim to reread the sentence from the beginning, from those regressions targeting the location where the potential error source has been expected. Also, the fine-grained analysis of the proportions suggests that the largest difference of targeting probability was found on the subject determiner. Thus, in anomalous sentences the determiner became the regression target more frequently than in non-anomalous sentences. This fits well to the role of the determiner in German because the determiner contains the case marking. In the SRA sentences used in our study, for example, an accusative determiner “den” would have solved the animacy violation. However, whereas this could be interpreted as more evidence for a linguistically constrained target selection, we again have to view these results with caution because the determiner was also the first word in the sentence. Thus, we have to leave it for future research to examine this pattern in more detail.

6.3.2 Results of the functional imaging data

The results of the functional imaging data revealed qualitatively different activation patterns for the onsets of the two types of eye movements, especially in regions associated with eye movement control and attention.

One of these regions is the medial SFG that is traditionally associated with visual and attentional processing. Henderson et al. (2015), for example, reported increased activations for text-reading vs. pseudo-text reading in the supplementary eye fields (SEF) and the supplementary motor area (SMA) which are both part of the medial SFG. In our analysis, by contrast, the activations in none of these regions were modulated by saccade type, rather, we found increased activations in the pre-supplementary motor area (pre-SMA; Picard & Strick, 1996) for regressive saccades compared to progressive saccades and baseline. The pre-SMA has often been related to control over voluntary actions, especially in the resolution of conflicts between motor plans (Garavan, Ross, Kaufman, & Stein, 2003; Nachev, Wydell, O'Neill, Husain, & Kennard, 2007; Nachev, Kennard, & Husain, 2008). Given that regressive eye movements are hypothesized to interrupt the automatized forward walk through the sentence, the pre-SMA might mediate between the initiation of these two eye movements.

We also found positive correlations with brain activity for both saccade types compared to baseline in parts of the right precentral sulcus on the border to the superior frontal sulcus (SFS), a region that is often referred to as the frontal eye fields (FEF; Grosbras, Laird, & Paus, 2005). Whereas the FEF play a crucial role in the execution of eye movements, the activity in the FEF does not seem to be modulated by saccade type, a finding that has also been reported for fixation duration by Henderson and colleagues (2015). This suggests that the FEF are not differentially involved in the interface between text comprehension and eye movement control. In contrast to the FEF, Henderson and colleagues claimed that the SEF may be "*an important component of the control network postulated by computational models of eye movement control in reading such as E-Z Reader, SWIFT, and CRISP that seek to account for fixation duration*" (p. 394). Since saccade generation is strongly coupled to the duration of a fixation, this raises the question as to why we did not find increased activations in the SEF in our experiment. An explanation might be that the SEF play a crucial role in the control of fixations (including linguistic computations) but not in the generation of the eye movement itself. This would make sense from the perspective that the initiation of an eye movement is the end, rather than the beginning of the linguistic processing. This assumption also fits with the lack of increased activations in areas associated with language processing as superior / middle temporal gyrus, which have also been reported by Henderson et al. for text reading vs. pseudo-text reading.

In addition to the qualitatively distinct activation pattern in the medial SFG, we found positive correlations with activation levels in parietal regions such as the IPS and the supramarginal gyrus (SMA) for regressions compared to progressions and baseline, which are also both assumed to be involved in eye movement control and visual attention (Grosbras et al., 2005; Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002). The IPS in particular has been linked to attention to motion direction (Corbetta & Shulman,

2002) as well as to processing serial order or language representations (Majerus et al., 2006). This is in accordance with the assumptions of the *Information Gathering Framework*, which claims that when making a regression, the representations of previous words are re-activated and confidence levels are re-computed in order to select the appropriate target position. In addition, the execution of a regressive eye movement also requires a mental representation of serial word order to find the selected target location in the sentence (Inhoff & Weger, 2005; Weger & Inhoff, 2007) and an attention shift to the left (Apel et al., 2012). Thus, the IPS may be involved in selecting a regression target and in directing the eyes to this location in the sentence.

Two more regions showed increased levels of BOLD activity for regressive saccades compared to progressive saccades and baseline, namely the dorso-lateral prefrontal cortex (DLPFC) and the ACC. Increased activation in the DLPFC has been found in a variety of tasks that require cognitive control such as decision making and inductive reasoning (Glenn, Raine, Schug, Young, & Hauser, 2009; Yang, Liang, Lu, Li, & Zhong, 2009). Crucially, the DLPFC has been found to be engaged in oculomotor tasks that include memory-guided saccades (Müri & Nyffeler, 2008; Pierrot-Deseilligny, Milea, & Müri, 2004). Given that the target selection of regressions (in contrast to progressions) requires the retrieval of word representations from the memory (at least partly), regressive eye movements can also be viewed as memory-guided saccades. Thus, the DLPFC might play a crucial role (together with the IPS) in the directing of regressive eye movements to a target that has to be retrieved from the memory. Henderson et al. (2015) also report increased activation in the DLPFC, but only in the left hemisphere and to a greater extent in pseudo-reading than text-reading. They argue that pseudo-reading involves more top-down controlled eye movements than is the case for the automatized saccades in natural reading, which can be applied to regressive saccades as well. In addition, however, they hypothesize that the lack of finding of right DLPFC activation may be due to the majority of right eye movements in their experiment. In our study, however, we found increased bilateral DLPFC activation for regressive eye movements only which speaks clearly against a correlation with the direction of the eye movement. Note also that the DLPFC also has direct connections to the pre-SMA but not the SMA (Lu, Preston, & Strick, 1994; Luppino, Matelli, Camarda, & Rizzolatti, 1993; Wang, Isoda, Matsuzaka, Shima, & Tanji, 2005), which is in line with our findings of increased activations in the pre-SMA but not in the SMA.

The positive correlations with brain activity in the ACC were partly driven by decreased activations for progressive eye movements compared to baseline. As the DLPFC, the ACC has also been hypothesized to be an important control component, namely engaged in conflict resolution in language, as for example proposed by Hagoort in the context of the MUC model (Hagoort, 2013; see also Ye & Zhou, 2009). Since saccades can be viewed as experiments that test hypotheses about the sensory input (Friston et al., 2012), regressions in particular can be viewed as a response to prediction error (see also the assumptions of the *Information Gathering Framework*). The ACC may be the area where these prediction errors are evaluated. For progressive saccades by contrast, no conflict arises and accordingly no increased activation in the ACC was observed.

In sum, the results of the fMRI data suggest that progressive and regressive saccades differ not only in the direction of the eye movement, but also that they reflect qualitatively distinct functional processes in reading. Whereas progressive saccades seem to be driven by a more automatized bottom-up mechanism requiring little cognitive control, regressive saccades seem to be driven top-down by prediction conflicts and go in hand with attention shifts. These findings are in line with current models of eye movement control such as SWIFT and the *Information Gathering Framework* that assume a default-like mechanism for forward saccades and a top-down control for backward saccades. Note, however, that we are not able to assess the role of linguistic processing for forward saccades in more detail. Despite increased activations in the right precentral gyrus compared to baseline which have often been reported in the context of phonological processing (Binder et al., 1997; Price, 2012), we found no hints for increased linguistic processing for progressive saccades. That said, we have to keep in mind that this may be due to the implicit baseline, which also included linguistic processing because it contained sentence reading.

In the second analysis of the fMRI data, we focused on the activations correlated with the length of an eye movement. For both types of saccades, we found that longer saccades were positively correlated with the level of BOLD activity in occipital regions in both hemispheres. Since these areas are involved in visual processing, these results suggest that the longer a saccade the more visual input is received, although it is assumed that the visual input received during a saccade cannot be used for higher-order processing (Rayner, 2009).

In addition, we found that for progressions, saccade length was correlated positively with bilateral activations in the middle-anterior part of the cingulate gyrus and sulcus (aMCC). The aMCC contains the cingulate motor areas, which in turn directly project to the motor cortices, and accordingly the aMCC has been hypothesized to play an important role in response selection (Vogt, 2005; Dum & Strick, 1991; Morecraft & Van Hoesen, 1992). Since it is assumed in eye movement models like SWIFT that longer progressive saccades occur due to cancelations of a previous eye movement that was generated in a highly-automated manner, we would expect that longer progressions are more likely to be driven by top-down control mechanisms than is the case for shorter saccades. This kind of eye movement control fits well to the role of the MCC in selecting motor responses.

For regressive eye movements by contrast, we found that saccade length was positively correlated with activations in the left DLPFC. As already discussed above, the DLPFC is assumed to be a control component that has been found to be engaged in memory-guided saccades (Müri & Nyffeler, 2008; Pierrot-Deseilligny et al., 2004). The *Information Gathering Framework* makes that claim that targets of longer regressive saccades (which are beyond the perceptual span) have to be retrieved from memory, at least their location in the sentence (see also Weger & Inhoff, 2007; Inhoff & Weger, 2005). From this perspective, increased activations in the left DLPFC are a very plausible finding. Note, however, that a positive correlation was only found in the left hemisphere, which accords with the results reported by Henderson et al. (2015; although it is still unclear what causes this lateralization).

In sum, these findings suggest that the length of an eye movement is not just subject to a random distribution but underlies higher-order control, especially for regressive saccades. Thus, the longer a regressive eye movement, the more top-down control is needed to find an appropriate target and to direct the eyes to this location in the sentence.

6.4 Summary

In Experiment 2 we examined the neural underpinnings of regressive eye movements during reading using concurrent fMRI / eye-tracking measures. We found that regressive and progressive saccades revealed qualitatively different activation patterns in the brain, indicating the different functional properties of the two types of saccades. Whereas progressive saccades seem to be driven bottom-up by a widely automatized saccade generation mechanism, regressive saccades are linked to increased cognitive control and attention shifts, possibly because they interrupt the default forward walk through the sentence due to a prediction error. Although the *Information Gathering Framework* was not developed to model the neural architecture of reading, the results of the combined fMRI / eye-tracking study fit well with the implicit assumptions of the framework, which postulates pronounced functional differences between the two types of saccades. Thus, the *Information Gathering Framework* also seems to provide an appropriate starting point to examine the neural correlates of natural reading, as well as developing a neural model of human reading behavior.

In addition to the fMRI study, we investigated the eye movement behavior when reading the German counterparts of the English SRA sentences. Since we also conducted a single eye-tracking study using the same experimental design as for the combined fMRI / eye-tracking study, we were able to demonstrate the reliability of the eye-tracking data recorded in the scanner. Furthermore, our results replicated some important findings of Experiment 1 which are directly predicted by the *Information Gathering Framework*, e.g. shorter fixation durations before regressions and two different groups of target positions for regressive eye movements. The results of Experiment 2 also replicated an important finding with regard to the language processing mechanisms, namely that the detection of an anomaly was widely unaffected by the high lexical association between arguments. But we were also able to detect crosslinguistic differences between English and German when processing SRA sentences, which emphasize once again the different role of information cues for the architecture of language processing within these two languages.

7. General Discussion

When humans use language, they can do so either by producing sounds or in writing letters. Both modalities employ the same language system of course, and enable people to share thoughts, feelings and information. Despite these similarities, however, auditory and visual linguistic input imposes different decoding demands. Whereas the auditory language has to be processed online without any control over the pace of the input, the processing of visual language usually allows for more flexibility. Given that the human language processing system generally works in a fascinatingly fast and accurate manner, we would expect that skilled readers learned to benefit from the specific characteristics of written language when they read (instead of applying the same linear processing mechanisms as for auditory language).

There are in particular two parameters that are crucial for reading and provide some room for individual control: The extension of fixation durations and the decision to reread. In the present thesis, we examined the role of regressive eye movements and their interaction with fixation durations during sentence reading in more detail. In the beginning, we hypothesized that regressions and fixation durations do not just sum up each other but have to be functionally distinguished: During a fixation, the reader can only deal with the currently available information, whereas a regressive eye movement allows for the intake of additional information, particularly of input to the left of the current fixation that has been (at least partly) processed earlier.

This idea also constitutes the basis of the *Information Gathering Framework* that forms the core of the present work. Within the *Information Gathering Framework,* fixation durations and regressions are assumed to be controlled by two independent mechanisms that are also sensitive to different linguistic levels (due to the time course of language processing), and the results of our empirical data indeed provide clear evidence for this functional distinction: Experiment 1 shows that regression rates can be manipulated independently of first pass fixation durations by using different comprehension questions. Experiment 2 on the other hand indicates that comprehension difficulties may increase the probability of regressions, but not necessarily lead to longer fixation durations in first pass reading at the same time. The observed neural correlates of regressive eye movements also differ – despite the activation correlated with the control of the eye movement per se – from the activation patterns reported for fixation durations by Henderson and colleagues (2015), which again suggests different functional underpinnings.

In addition, in both studies we replicated the finding that fixation durations tend to be shorter before regressive saccades compared to progressive saccades, a result that is directly predicted from the architecture of the *Information Gathering Framework*.

Besides the functional distinction between fixation durations and regressions, the *Information Gathering Framework* also proposes that two different types of regressive saccades can be distinguished. Whereas the goal of all regressive saccades is assumed to gather additional information relevant in the course of sentence interpretation, regressions can either be caused by integration difficulties due to prediction errors or by missing expected evidence without integration difficulties. The latter group of regressions that occur particularly at the end of a sentence has never been the focus of broader research nor formed part of eye movement control models to this point. Our data, however, provides strong empirical evidence across both studies for the dissociation of these two types of regressive saccades: Whereas the question difficulty manipulation of Experiment 1 did clearly affect the regression probability from the end of a sentence, it essentially had no impact on regressions occurring earlier in the sentence, which are assumed to be caused by integration difficulties. Also, for the first time we reported differences in fixation durations before the two types of regressions, which were directly predicted from the *Information Gathering Framework*. We are of the view that this distinction between regressive saccade types within the *Information Gathering Framework* provides an excellent starting point to explore the role of regressive eye movements in more detail through future research.

In both experiments we were also able to shed more light on the target positions of regressive eye movements. The *Information Gathering Framework* again predicts that two types of target selection mechanisms can be distinguished: The first mechanism computes the target within the perceptual span on the basis of clear linguistic computations and therefore claims a strong linguistic guidance on regression landing sites (as e.g. proposed by Frazier & Rayner, 1982). The second mechanism which is applied if mechanism 1 does not deliver a clear result, works on the basis of experience based strategies (c.f. von der Malsburg & Vasishth, 2011, 2013). Although our experiments were not designed to directly test these assumptions, we were nonetheless able to find hints for this distinction across both studies by identifying two groups of landing positions. There were also hints that there exists an invisible boundary for the target selection at about 15-18 characters to the left of the current fixation; after that the number of regression targets clearly decreases. This boundary may indicate the border of the perceptual span which is shifted to the left when making a regressive eye movement (Apel et al., 2012). However, more research is needed to examine the target selection of regressive eye movements in more detail and to incorporate these results into the architecture of the *Information Gathering Framework*.

In sum, the *Information Gathering Framework* as presented is the first model of eye movement control that explicitly focuses on the interplay between fixation durations and regressive eye movements during reading, that utilizes human eye movement data and that can be applied to all types of regressive saccades. The described framework, however, is not without its limitations. The results of our studies point to several aspects and outstanding problems that have to be examined in more detail through future research.

One of the most important goals for future research would be to implement the framework in a computational model that allows for simulations of human reading behavior as well. Whereas this is a very challenging task by itself, it would probably require a better formalization of the underlying language model. Currently the *Information Gathering Framework* assumes that new input is matched with the predictions from former language material on the basis of explicit production rules. This is, however, a very vague description and it remains unclear how these rules are defined and how they interact with other information sources like world knowledge. For example, it is still an open question as to how the language model proposed here may account for the detailed processing of SRA sentences that were the focus of the present experiments.

Although we criticized the current models of eye movement control because of their underspecified assumptions about the underlying language processing mechanisms, we must acknowledge that the language model we proposed within the *Information Gathering Framework* is also a simplification which in all probability cannot account for the complexity of language processing. This again points to the very crucial aspect that we have already discussed in the Introduction, namely, that regressive eye movements always have to be viewed in the interplay between oculomotor processes on the one hand, and language processing mechanisms on the other. Thus, an understanding of the role of regressive eye movements during reading also requires a deep understanding of the way the linguistic information is decoded, including higher levels of language processing.

Another crucial issue for future research is the question concerning the precise nature of the backward regression strategies. We have already mentioned that these strategies are assumed to rely on language experience and should be effective, which means that they lead to a high degree of accuracy that is reached within a short time. Also there should exist only a limited number of strategies, in response to a broader task category each. One example for such a strategy could be the forward strategy that moves the eyes back to the beginning of the sentence in order to start rereading. Since the results of our two experiments which have been reported also show a clear tendency to regress to the beginning of the sentence, this might indeed be a very useful strategy. However, these considerations also imply that it would be worthwhile to take into account regression patterns instead of single targets, as proposed by von der Malsburg and Vasishth (2011; 2013).

Another aspect of backward regression strategies might be their development over a lifetime. Since these strategies are assumed to rely on language experience and the language experience increases over a lifetime, it should also be reasonable to expect that the strategies applied by adult readers differ from those applied by children (despite the more careful first pass strategy for developing readers that has been discussed above).

First of all, we have to keep in mind that children in their first years of reading development show a higher variability between readers with regard to reading behavior (fixation durations, regression rates and so on) than adult readers do (McConkie et al., 1991). Thus, we also would expect a high variability of reading strategies between children. Nonetheless, children possibly apply a strategy that needs less linguistic knowledge, so that they are less likely to adopt a language specific strategy as the one proposed above. On the other hand, there exists evidence that the physical properties underlying the control of eye movements are not fully developed when children begin learning to read (Luna, Velanova, & Geier, 2008; Seassau & Bucci, 2013). Thus, saccades in younger children are imprecise which could lead to a lower performance. Accordingly, children should adopt a strategy that does not require long saccades. Perhaps they tend to reread the previous word when faced with difficulties. Again, this has to be the subject of future research.

In addition to developments in the control of eye movements, there is evidence that the perceptual span differs between young and old adults, with an age-related reduction of the perceptual span to the right of the current fixation (Rayner, Castelhano, & Yang, 2010; Risse & Kliegl, 2011). Given the importance of the perceptual span for the regression target selection within the *Information Gathering Framework*, this may also lead to differences in the landing site distributions of regressive eye movements between young and older readers. However, more research is needed here, especially on the size of the perceptual span when making a regressive eye movement.

Another crucial problem of *the Information Gathering Framework*, however, is its time course of oculomotor processing and saccade planning. As already mentioned in chapter 1, saccades are motor responses that need time to be planned and executed (Rayner, 1998; 2009). Accordingly, the execution of a regressive eye movements does not immediately follow the detection of difficulties, but is delayed. In addition, it has been argued that due to the time interval that is needed for initiating a saccade (about 175–200 ms), the time of the current fixation is (at least in some cases) too short to plan a motor response such as an eye movement during this fixation (see Vitu, 2011, for a recent discussion).

The *Information Gathering Framework*, however, does not account for saccade latency so far. In particular, it is assumed that the eye movement is triggered and executed as soon as the confidence level crosses a certain threshold. This is, of course, a very unrealistic assumption from the motor response perspective and thus the question arises as to how we might incorporate saccade latency into the current model without adding excessive complexity.

One opportunity would be to assume that a saccade is triggered but not executed as soon as the confidence level is crossed. In this case, the eyes still fixate word *n* although a saccade to word *n+1* has already been triggered, which would allow for a further computation of confidence levels of word *n* during a fixation on word *n* after reaching the forward threshold. In particular, this means that the confidence level of word *n* could fall under the forward threshold even during a fixation of word *n* (which is, by contrast, not possible within the current architecture of the model). If this happens, a progressive saccade is still executed to word *n+1* because the motor program (working in a highly automatic manner for progressive saccades) cannot be canceled (at least after a certain amount of time, as for example assumed by the SWIFT model). But in addition, a regressive saccade is triggered (because the confidence level fell under the forward threshold), which would cause a regressive eye movement that is executed shortly after the eyes moved to word *n+1*. This again would for instance explain why fixations before regressive eye movements tend to be shorter.

However, more research is needed to integrate saccade latency into the current model and to see whether it still accounts for the results reported in the literature and our two experiments presented above.

Another problem is the question of how spatial information can be incorporated into the *Information Gathering Framework*. Since we assume that the target of a regression is not subject to a random or fixed mechanism but rather is selected on the basis of linguistic computations, we consequently have to conclude that this would only make sense if the eyes are also able to find this particular target in the sentence. Specifically, the problem arises for targets that are beyond the currently visible area of the text. Thus, the execution of a regressive eye movement needs some information about the location of the target in order for the eyes to move there.

This information can either be coded on a spatial sentence map where each word is marked with a certain identification measure, or it could form part of the lexical representation instead (Inhoff & Weger, 2005; Weger & Inhoff, 2007). Both concepts, however, reveal some problems because the storage and retrieval of spatial information requires memory capacities which are hardly in line with current findings (e.g. the representation of word order, see McElree, 2006). Thus, more research is required to incorporate the storage and retrieval of spatial information into the *Information Gathering Framework*.

In addition, the role of visual information that guides the eyes' landing position is still unclear. Whereas Mitchell and colleagues (2008) report some layout effects so that regressions rarely target a word in the previous line, the visual information cannot be a general determinant of the target selection due to the restrictions of the oculomotor system. Nonetheless, visual information could play a role in planning and executing a regressive eye movement. It would be interesting, for example, to see what happens if the visual salience of words is increased by using colors or certain fonts. Maybe this allows for a better and more accurate retrieval of this word from the memory, which leads to an increased probability that this word becomes the target of regression.

It is also possible that visual salience does not affect the target selection of regressions at all. An interesting issue for future research is therefore to examine these influences in more detail. Furthermore, the detailed implementation in the *Information Gathering Framework* still remains unclear.

The results of the studies reported above also point to another issue that is crucial for all eye movement control models during reading, namely, the question of the underlying linguistic units. Specifically, the *Information Gathering Framework* focuses on words rather than phrases. We have done so because most current models of eye movement control (like SWIFT and E-Z Reader) are word-based accounts which follow from the assumption that the identification of words is the crucial task in reading (cf. for example Bicknell & Levy, 2012).

This assumption, however, is not necessarily correct because reading is not only the extraction of visual information, but rather the gathering of sentence and text comprehension. Thus, the most important factor in reading is in fact the assignment of linguistic functions to visual input. The identification of words is a crucial prerequisite for this task, of course, but it is not the top of the flagpole. In particular, we have to keep in mind that words appear in the context of other words and can be combined into phrases. These phrases have a function as a whole. Consider, for example, the complex nominal phrase *"the little girl from New York"* which can either be the actor *("The little girl from New York is writing a book"*) or the undergoer[35] of an action *("The burglar was firing at the little girl from New York*"). It is thus plausible to assume that eye movements in reading can be better explained by approaches that are focusing on phrases rather than words (c.f. also the results of Experiment 2, which showed differences in single fixation times but not in phrase-based measures between the two experiment types). This is especially true for models that are focusing on regressive eye movements which are assumed to be driven by higher-order linguistic processing.

In the tradition of research on sentence processing it is this view which has prevailed, so that most studies on sentence processing report reading time measures for phrases rather than words. We therefore think that an important issue for future research is to determine in how far regressive eye movements are based on words or phrases instead, and how these considerations can be incorporated into the current model.

Finally, it seems to us that an important implication for future research is the role of regressive eye movements in developing an efficient reading strategy; in particular for

[35] If we assume that there exist just these two broad functions (actor and undergoer) for arguments in sentences.

pupils in school and young adults. Since written language is a substantial part of our society, the ability to read is a prerequisite for all fields of higher education. However, reading and especially the comprehension of higher demanding texts is a task with which many students struggle. Accordingly, we should raise the question as to how our research on the cognitive (and neural) mechanisms of reading might help us to develop educational settings that improve reading abilities. Whereas research in this area focuses mainly on reading beginners, systematic empirical evidence on advanced readers' behavior is rare. Most importantly, skilled reading does not just require oculomotor processes but rather a good language skill. Or, as Rayner and colleagues express it in the context of speed-reading techniques:

> *The kind of practice that will help reading is practice that helps people to identify words and comprehend better, not just take in visual information faster. [...] The practice that is required to become a better reader is thus practice with language. (Rayner et al., 2016, p. 28f.)*

Language is thus the most important parameter in reading. Nonetheless, there are different ways as to how the language skill can be practiced. In this thesis, we have also outlined how regressive eye movements interact with (higher-order) language skills. Therefore we suggest that regressive eye movements could be used as a diagnostic tool to identify difficulties in language processing and further, to develop appropriate educational settings that practice relevant language structures and help students to comprehend demanding linguistic input. The characteristics of written language and especially of regressive eye movements in particular seem to be well suited to practice language skill.

8. Conclusion

In conclusion, we want to again emphasize the crucial role of regressive eye movements during reading: Whereas rereading has often been viewed as some kind of obstacle encountered on the way to fast reading behavior (especially in the context of speed reading techniques, see Rayner et al., 2016, for a discussion), recent research seems to indicate that the fine-tuned interplay between fixation durations and rereading in particular enables readers to be both fast and accurate at the same time (Bicknell & Levy, 2010). Thus, the use of regressive eye movements cannot be attributed to a low reading skill per se, but rather is one of the substantial characteristics of a skilled reader. This in turn means that regressions are not a waste of time – they rather save time.

Bibliography

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 529.

Altmann, G. T., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*(5), 685–712.

Andriessen, J. J. & A. H. De Voogd. (1973). Analysis of eye movement patterns in silen reading. *IPO Annual Progress Report*, *8*, 29–34.

Apel, J. K., Henderson, J. M., & Ferreira, F. (2012). Targeting regressions: Do readers pay attention to the left? *Psychonomic Bulletin & Review*, *19*(6), 1108–1113. https://doi.org/10.3758/s13423-012-0291-1

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Bader, M. (2015). Leseverstehen und Sprachverarbeitung. In *Lesen. Ein interdisziplinäres Handbuch.* (pp. 141–168). Berlin / Boston: de Gruyter.

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*. Retrieved from http://arxiv.org/abs/1406.5823

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*, 967–983.

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics* (pp. 1168–1178). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1858800

Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: A statistical analysis. In *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 931–936). Retrieved from https://mindmodeling.org/cogsci2011/papers/0210/paper0210.pdf

Bicknell, K., & Levy, R. (2012). The utility of modelling word identification from visual input within models of eye movements in reading. *Visual Cognition*, *20*(4–5), 422–456. https://doi.org/10.1080/13506285.2012.668144

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17*(1), 353–362.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The Role of Prominence Information in the Real-Time Comprehension of Transitive Constructions: A Cross-Linguistic Approach. *Language and Linguistics Compass*, *3*(1), 19–58. https://doi.org/10.1111/j.1749-818X.2008.00099.x

Bouma, H. (1978). Visual Search and Reading: Eye Movements and Functional Visual Field: A Tutorial Review. In *Attention and performance VII: Proceedings of the seventh international symposium on attention and performance, Sénaque, France, August1-6, 1976* (pp. 115–147). London: Erlbaum.

Bouma, H., & De Voogd, A. H. (1974). On the control of eye saccades in reading. *Vision Research*, *14*, 273–284.

Buswell, G. T. (1922). *Fundamental reading habits, a study of their development*. Chicago: Chicago University Press.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *The Quarterly Journal of Experimental Psychology*, *70*(7), 1380–1405. https://doi.org/10.1080/17470218.2016.1186200

Corbetta, M., & Shulman, G. L. (2002). CONTROL OF GOAL-DIRECTED AND STIMULUS-DRIVEN ATTENTION IN THE BRAIN. *Nature Reviews Neuroscience*, *3*(3), 215–229. https://doi.org/10.1038/nrn755

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). fMRI Clustering in AFNI: False-Positive Rates Redux. *Brain Connectivity*, *7*(3), 152–171.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dum, R. P., & Strick, P. L. (1991). The origin of corticospinal projections from the premotor areas in the frontal lobe. *Journal of Neuroscience*, *11*(3), 667–689.

Dürscheid, C. (2012). *Einführung in die Schriftlinguistik* (4., überarbeitete und aktualisierte Auflage). Göttingen (Germany): Vandenhoeck & Ruprecht.

Eichenbaum, H. (2013). Erratum: Memory on time. *Trends in Cognitive Sciences*, *17*(5), 255. https://doi.org/10.1016/j.tics.2013.04.002

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources? *Journal of Memory and Language*, *54*(4), 541–553. https://doi.org/10.1016/j.jml.2005.12.006

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, *4*(3–4), SI155–SI209.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23*(5), 407–434.

Forster, K. I., & Chambers, S. M. (1973). Lexical Access and Naming Time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635.

Frazier, L., & Rayner, K. (1982). Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology*, *14*, 178–210.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*(1), 45–63.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*(1), 136–158. https://doi.org/10.1016/j.jml.2004.03.008

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581–590.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00151

Garavan, H., Ross, T. ., Kaufman, J., & Stein, E. . (2003). A midline dissociation between error-processing and response-conflict monitoring. *NeuroImage*, *20*(2), 1132–1139. https://doi.org/10.1016/S1053-8119(03)00334-3

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Glenn, A. L., Raine, A., Schug, R. A., Young, L., & Hauser, M. (2009). Increased DLPFC activity during moral decision-making in psychopathy. *Molecular Psychiatry*, *14*(10), 909–911. https://doi.org/10.1038/mp.2009.76

Grosbras, M.-H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, *25*(1), 140–154. https://doi.org/10.1002/hbm.20145

Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, *15*(5), 643–657.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00416

Henderson, J. M., Choi, W., Luke, S. G., & Desai, R. H. (2015). Neural correlates of fixation duration in natural reading: Evidence from fixation-related fMRI. *NeuroImage*, *119*, 390–397. https://doi.org/10.1016/j.neuroimage.2015.06.072

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1390–1400. https://doi.org/10.1037/a0036330

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443. https://doi.org/10.1016/j.jml.2005.12.001

Inhoff, A. W., & Weger, U. W. (2005). Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory & Cognition*, *33*(3), 447–461.

Ishida, T., & Ikeda, M. (1989). Temporal properties of information extraction in reading studied by a text-mask replacement technique. *Journal of the Optical Society of America*, *6*(10), 1624–1632.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In *New methods in reading comprehension research* (pp. 151–182). Hillsdale, NJ, US: Erlbaum.

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1561–1566. https://doi.org/10.1037/a0020693

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

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.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the Mind During Reading: The Influence of Past, Present, and Future Words on Fixation Durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In *Reanalysis in sentence processing* (pp. 247–285). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-015-9070-9_8

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, S., & Phillips, C. (2015). Aligning Grammatical Theories and Language Processing Models. *Journal of Psycholinguistic Research*, *44*(1), 27–46. https://doi.org/10.1007/s10936-014-9329-z

Lu, M.-T., Preston, J. B., & Strick, P. L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *Journal of Comparative Neurology*, *341*(3), 375–392.

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. https://doi.org/10.1016/j.bandc.2008.08.019

Luppino, G., Matelli, M., Camarda, R., & Rizzolatti, G. (1993). Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *Journal of Comparative Neurology*, *338*(1), 114–140.

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. https://doi.org/10.1037//0033-295X.109.1.35

MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, *3*, 711. https://doi.org/10.1038/ncomms1715

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150.

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., … Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *NeuroImage*, *32*(2), 880–891. https://doi.org/10.1016/j.neuroimage.2006.03.048

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In *ACL-MIT Press series in natrual language processing. Cognitive models of speech processing: Psycholinguistic and computational perspectives.* (pp. 148–172). Cambridge, MA: MIT Press.

Marsman, J. B. C., Renken, R., Velichkovsky, B. M., Hooymans, J. M. M., & Cornelissen, F. W. (2012). Fixation based event-related fmri analysis: Using eye fixations as events in functional magnetic resonance imaging to reveal cortical processing during the free exploration of visual images. *Human Brain Mapping*, *33*(2), 307–318. https://doi.org/10.1002/hbm.21211

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological Bulletin*, *81*(12), 899.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixation on words. *Vision Research*, *28*(10), 1107–1118.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Attention, Perception, & Psychophysics*, *46*(3), 245–253.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578–586.

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*, 365–368.

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wolff, P. M. (1991). Children's eye movement during reading. In *Vision and visual dyslexia* (pp. 251–262). London: Macmillan Press.

McDonald, S. A., & Shillcock, R. C. (2004). The potential contribution of preplanned refixations to the preferred viewing location. *Attention, Perception, & Psychophysics*, *66*(6), 1033–1044.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, *46*, 155–200.

McElree, B. (2015). Memory Processes Underlying Real-Time Language Comprehension. In *Remembering: Attributions, Processes, and Control in Human Memory* (pp. 133–152). New York and London: Psychology Press.

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*(4), 551–561.

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2016). The Importance of Reading Naturally: Evidence from Combined Recordings of Eye Movements and Electric Brain Potentials. *Cognitive Science*.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6), 721–728.

Miller, S. D., & Smith, D. E. P. (1989). Relations among oral reading, silent reading and listening comprehension of students at differing competency levels. *Reading Research and Instruction*, *29*(2), 73–84.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*(3), 266–293. https://doi.org/10.1016/j.jml.2008.06.002

Morecraft, R. J., & Van Hoesen, G. W. (1992). Cingulate Input to the Primary and Supplementary Motor Cortices in the Rhesus Monkey: Evidence for Somatotopy in Areas 24c and 23c. *The Journal of Comparative Neurology*, *322*, 471–489.

Müri, R. M., & Nyffeler, T. (2008). Neurophysiology and neuroanatomy of reflexive and volitional saccades as revealed by lesion studies with neurological patients and transcranial magnetic stimulation (TMS). *Brain and Cognition*, *68*(3), 284–292. https://doi.org/10.1016/j.bandc.2008.08.018

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856–869. https://doi.org/10.1038/nrn2478

Nachev, P., Wydell, H., O'Neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. *NeuroImage*, *36*, T155–T163. https://doi.org/10.1016/j.neuroimage.2007.03.034

Newell, A. (1973). Production systems: Models of control structures. In *Visual information processing*. New York: Academic Press.

Oldfield, R. C. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113.

O'Regan, J. K. (1990). Eye movements in reading. *Reviews of Oculomotor Research*, *4*, 395–453.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. *Eye Movements and Visual Cognition*, 333–354.

O'Regan, J. K., & Lévy-Schoen, A. (1987). Eye-movements strategy and tactics in word recognition and reading. In *Attention and performance 12: The pychology of reading* (pp. 363–383). Hillsdale, NJ, US: Lawrence Erlbaum Assoc.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex*, *6*(3), 342–353.

Pierrot-Deseilligny, C., Milea, D., & Müri, R. M. (2004). Eye movement control by the cerebral cortext. *Current Opinion in Neurology*, *17*, 17–25.

Pollatsek, A., & Rayner, K. (1990). Eye Movements and Lexical Access in Reading. In *Comprehension processes in reading* (pp. 143–163).

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Rayner, K. (1978). Eye movement latencies for parafoveally presented words. *Bulletin of the Psychonomic Society*, *11*(1), 13–16.

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education*, *6*(6), 53–60.

Rayner, K. (1986). Eye Movements and the Perceptual Span in Beginning and Skilled Readers. *Journal of Experimental Child Psychology*, *41*, 211–236.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*(3–4), 242–258. https://doi.org/10.1080/13506285.2013.879084

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*, 468–469.

Rayner, K., Carlson, M., & Frazier, L. (1983). The Interaction of Syntax and Semantics During Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.

Rayner, K., Castelhano, M. S., & Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging*, *25*(3), 714–718. https://doi.org/10.1037/a0019199

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*, 191–201.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology Section A*, *53*(4), 1061–1080. https://doi.org/10.1080/713755934

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726. https://doi.org/10.1016/j.visres.2007.05.007

Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In *The Oxford handbook of eye movements* (pp. 751–766). Oxford University Press.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, *14*(4), 385–388.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Pergamon Press*, *16*, 829–837.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. London: Prentice-Hall.

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So Much to Read, So Little Time How Do We Read, and Can Speed Reading Help? *Psychological Science in the Public Interest*, *17*(1), 4–34.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes [special issue]. *Language and Cognition Processes*, *4*, 21–49.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188.

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 912.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The Effect of Plausibility on Eye Movements in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*, 537–544.

Rayner, K., Yang, J., Castelhano, M. S., & Liversedge, S. P. (2011). Eye movements of older and younger readers when reading disappearing text. *Psychology and Aging*, *26*(1), 214–223. https://doi.org/10.1037/a0021279

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4–22. https://doi.org/10.1016/j.cogsys.2005.07.002

Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check: E-Z Reader and event-related potentials. *Psychophysiology*, *48*(7), 993–1003. https://doi.org/10.1111/j.1469-8986.2011.01169.x

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21. https://doi.org/10.3758/PBR.16.1.1

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*(2), 177–206. https://doi.org/10.1016/j.cogpsych.2012.03.001

Richlan, F., Gagl, B., Hawelka, S., Braun, M., Schurz, M., M. Kronbichler, & Hutzler, F. (2014). Fixation-Related fMRI Analysis in the Domain of Reading Research: Using Self-Paced Eye Movements as Markers for Hemodynamic Brain Responses During Visual Letter String Processing. *Cerebral Cortex*, *24*(10), 2647–2656. https://doi.org/10.1093/cercor/bht117

Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging*, *26*(2), 451.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't Believe What You Read (Only Once): Comprehension Is Supported by Regressions During Reading. *Psychological Science*, *25*(6), 1218–1226. https://doi.org/10.1177/0956797614531148

Schutz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 9–9. https://doi.org/10.1167/11.5.9

Seassau, M., & Bucci, M.-P. (2013). Reading and visual search: a developmental study in normal children. *PloS One*, *8*(7), e70261.

Seidenberg, M. S. (2012). Computational models of reading. *The Cambridge Handbook of Psycholinguistics*, 186.

Simon, O., Mangin, J.-F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*(3), 475–487.

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology*, *27*(5), 511–530. https://doi.org/10.1080/20445911.2014.993990

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569. https://doi.org/10.1037/0278-7393.33.3.550

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation: Predictability and Eye Movements. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Taylor, S. E., Frackenpohl, H., & Petee, J. L. (1960). Grade level norms for the components of the fundamental reading skill. In *Research and Information Bulletin (Vol. 3)*. Huntington, NY: Educational Developmental Labs, Inc.

Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension: Interference in Comprehension. *Language and Linguistics Compass*, *6*(4), 193–211. https://doi.org/10.1002/lnc3.330

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, *67*(4), 703–727. https://doi.org/10.1080/17470218.2013.826258

Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In *The Oxford Handbook of Eye Movements* (pp. 731–749). Oxford University Press.

Vitu, F., & McConkie, G. W. (2000). Regressive Saccades and Word Perception in Adult Reading. In *Reading as a Perceptual Process* (pp. 301–326). Elsevier Science Ltd.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nature Reviews Neuroscience*, *6*(7), 533–544. https://doi.org/10.1038/nrn1704

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*(2), 109–127. https://doi.org/10.1016/j.jml.2011.02.004

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. https://doi.org/10.1080/01690965.2012.728232

Wang, Y., Isoda, M., Matsuzaka, Y., Shima, K., & Tanji, J. (2005). Prefrontal cortical cells projecting to the supplementary eye field and presupplementary motor area in the monkey. *Neuroscience Research*, *53*(1), 1–7. https://doi.org/10.1016/j.neures.2005.05.005

Waters, G. S. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 51–79. https://doi.org/10.1080/713755607

Waters, G. S., & Caplan, D. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 51–79. https://doi.org/10.1080/713755607

Weekes, B. S. (1997). Differential Effects of Number of Letters on Word and Nonword Naming Latency. *The Quarterly Journal of Experimental Psychology Section A*, *50*(2), 439–456. https://doi.org/10.1080/713755710

Weger, U. W., & Inhoff, A. W. (2007). Long-range regressions to previously read words are guided by spatial and verbal memory. *Memory & Cognition*, *35*(6), 1293–1306.

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2017). Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, 1–37. https://doi.org/10.1080/17470218.2017.1307862

Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology*, *66*(3), 548–562. https://doi.org/10.1080/17470218.2011.625094

Yang, S. (2006). An oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research*, *7*(1), 56–69. https://doi.org/10.1016/j.cogsys.2005.07.005

Yang, S. -N., & McConkie, G. W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, *41*, 3567–3585.

Yang, S. -N., & McConkie, G. W. (2004). Saccade generation during reading: Are words necessary? *European Journal of Cognitive Psychology*, *16*(1–2), 226–261. https://doi.org/10.1080/09541440340000231

Yang, Y., Liang, P., Lu, S., Li, K., & Zhong, N. (2009). The role of the DLPFC in inductive reasoning of MCI patients and normal agings: An fMRI study. *Science in China Series C: Life Sciences*, *52*(8), 789–795. https://doi.org/10.1007/s11427-009-0089-1

Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, *48*(1), 280–290. https://doi.org/10.1016/j.neuroimage.2009.06.032

ZUSAMMENFASSUNG IN DEUTSCHER SPRACHE

Wenn wir lesen, gleiten unsere Augen nicht durch den Satz, sondern sie springen. Während aber die meisten dieser Sprünge, auch Sakkaden genannt, in der Leserichtung erfolgen, bewegen sich bei ca. 15% der Sakkaden die Augen in die entgegengesetzte Richtung. Obwohl diese so genannten Regressionen der Leseforschung seit Jahrzehnten wohlbekannt sind, bilden ihre genauen Ursachen und Funktionen nach wie vor ein Forschungsdesiderat, was auch darin begründet liegt, dass eine Erklärung sowohl okulomotorische als auch linguistische Aspekte berücksichtigen muss.

Die vorliegende Arbeit versucht, sich diesem Problem zu nähern. Dazu wird im ersten Teil nach einer kritischen Analyse der aktuellen Forschungslage ein eigenes Blickbewegungsmodell entworfen, das so genannte *Information Gathering Framework*, welches das komplexe Zusammenspiel von Fixationszeiten und Regressionsraten während des Lesens zu erklären versucht. Das Modell basiert auf der einfachen Annahme, dass eine Regression immer dann ausgelöst wird, wenn zusätzliche Information über vorherigen Input benötigt wird, genauer gesagt, Information über die Identität eines vorangegangen Wortes.

Im zweiten Teil der Arbeit werden zwei empirische Studien vorgestellt, die wichtige Annahmen des entwickelten Modells testen und weitere Erkenntnisse über die Rolle von regressiven Blickbewegungen für die Satzinterpretation liefern. In der ersten Studie, einer Blickbewegungsstudie, wurde der Einfluss von unterschiedlichen Verständnisfragen auf das Regressionsmuster beim Lesen von normalen und anormalen englischen Sätzen untersucht. In der zweiten Studie lag der Fokus neben einer sprachvergleichenden Untersuchung mit deutschen Sätzen auf den neurobiologischen Unterschieden zwischen progressiven und regressiven Sakkaden. Dazu wurden Blickbewegungsmessungen in ein Verfahren mit funktionalen Bildgebung (fMRT) integriert und die neuronale Antwort des Gehirns auf die beiden Sakkadentypen analysiert.

Die Ergebnisse beider Studien bestätigen wichtige Annahmen des *Information Gathering Frameworks*. So konnte im Besonderen eine klare funktionale Dissoziation zwischen verlängerten Fixationszeiten und Regressionsraten, die beide als Indiz für linguistische Verarbeitungsschwierigkeiten gelten, nachgewiesen werden. Darüber hinaus wurde auch die Annahme des Modells, dass neben der gemeinsamen Grundfunktion zwei unterschiedliche Typen von Regression existieren, anhand der Blickbewegungsmuster bestätigt. Die neuronalen Ergebnisse liefern ferner weitere Evidenz für die funktionale Dissoziation zwischen progressiven und regressiven Sakkaden.

Zusammengefasst präsentiert die vorliegende Arbeit das erste umfassende kognitive Blickbewegungsmodell zu regressiven Sakkaden während des Lesens, das den hierarchischen linguistischen Verarbeitungsprozess berücksichtigt und gleichzeitig die wichtigsten Befunde des menschlichen Leseverhaltens erklären kann.

Beschreibung des Eigenanteils

Während der erste Teil dieser Arbeit, der auch die Entwicklung des *Information Gathering Frameworks* beinhaltet, vollständig meine eigene Leistung ist (von konstruktiven Vorschlägen und Diskussionen abgesehen), so entstanden die beiden Experimente im zweiten Teil in Zusammenarbeit mit verschiedenen Kolleginnen und Kollegen. Auch wenn dieses Vorgehen der gängigen Forschungspraxis in der Psycho- und Neurolinguistik entspricht, möchte ich an dieser Stelle kurz erläutern, welche Teile ich hier selbst geleistet habe.

Bei Experiment 1 habe ich die Stimuli vollständig selbst konstruiert und vorgetestet, die technische Implementation vorgenommen und einen Großteil der Probanden gemessen. Die Analyse inklusive Vorverarbeitung der Daten ist ebenfalls meine eigene Leistung sowie große Teile der Interpretation und des Verfassens sowie Vorbereitens des Manuskripts für die Veröffentlichung (welches auch in Teilen in der vorliegenden Arbeit übernommen wurde).

Bei Experiment 2 habe ich ebenfalls alle Schritte von der Konstruktion der Stimuli über die technische Vorbereitung und Durchführung der beiden Teilexperimente sowie deren Auswertung und die Verfassung des Manuskriptes selbst vorgenommen. Allerdings erhielt ich Unterstützung sowohl bei der Entwicklung des Designs als auch bei der technischen Vorbereitung und Durchführung sowie der Auswertung. Auch Teile des Manuskriptes wurden von den Co-Autoren überarbeitet.

Für alle Teile der vorliegenden Arbeit, die sich auf das *Information Gathering Framework* beziehen (sowohl im theoretischen als auch im experimentellen Teil, inklusive der Hypothesenentwicklung und –testung) wurde keinerlei zusätzliche Unterstützung in Anspruch genommen (außer der im ersten Abschnitt genannten).

ERKLÄRUNG ZUR EIGENSTÄNDIGEN ABFASSUNG

Hiermit versichere ich, dass ich die vorgelegte Dissertation mit dem Titel

The Information Gathering Framework.
A Cognitive Model of Regressive Eye Movements during Reading.

selbst und ohne weitere als die genannte Hilfe verfasst, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt (einschließlich des World Wide Web und anderen elektronischen Text- und Datensammlungen), alle vollständig oder sinngemäß übernommene Zitate als solche gekennzeichnet sowie die Dissertation in der vorliegenden oder einer ähnlichen Form noch keiner in- oder ausländischen Hochschule anlässlich eines Promotionsgesuches oder zu anderen Prüfungszwecken eingereicht habe.

Marburg, den 03.10.2017

Anna Fiona Weiß

CURRICULUM VITAE BREVE

Anna Fiona Weiß

2008	Erwerb der Allgemeinen Hochschulreife (Abitur)
2013	1. Staatsexamen für das Lehramt an Gymnasien mit den Fächern Latein, Deutsch und Erziehungswissenschaften, Philipps-Universität Marburg
seit 2013	Promotionsstudium im Fach Linguistik, Philipps-Universität Marburg, Betreuerin: Prof. Dr. Ina Bornkessel-Schlesewsky (ehemals Universität Marburg, seit 2014 an der University of South Australia, Adelaide)

APPENDIX

1. Sentence material used in Experiment 1

a) Semantic Reversal Anomalies

	N = non-anomalous, A = anomalous, H = highly associated, L = low associated
AH1	Currently the book is writing the author after the last one became a best seller.
AL1	Currently the book is fixing the author after the editor's comments.
NH1	Currently the author is writing the book after the last one became a best seller.
NL1	Currently the author is fixing the book after the editor's comments.
AH2	During the game the ball is throwing the quarterback with great skill.
AL2	During the game the ball is finding the quarterback difficult to throw.
NH2	During the game the quarterback is throwing the ball with great skill.
NL2	During the game the quarterback is finding the ball difficult to throw.
AH3	In the spring the tree is planting the ranger that was imported from Japan.
AL3	In the spring the tree is sawing the ranger to check its bark.
NH3	In the spring the ranger is planting the tree that was imported from Japan.
NL3	In the spring the ranger is sawing the tree to check its bark.
AH4	This Friday evening the wine is serving the waiter while the musician starts playing.
AL4	This Friday evening the wine is forcing the waiter on the wealthy guests.
NH4	This Friday evening the waiter is serving the wine while the musician starts playing.
NL4	This Friday evening the waiter is forcing the wine on the wealthy guests.
AH5	After breakfast the wheelchair is pushing the nurse to the bathroom.
AL5	After breakfast the wheelchair is removing the nurse from the bathroom.
NH5	After breakfast the nurse is pushing the wheelchair to the bathroom.
NL5	After breakfast the nurse is removing the wheelchair from the bathroom.
AH6	As the sun sets the treasure is burying the pirate in the sand.
AL6	As the sun sets the treasure is hugging the pirate with great joy.
NH6	As the sun sets the pirate is burying the treasure in the sand.
NL6	As the sun sets the pirate is hugging the treasure with great joy.
AH7	This time the crown is wearing the princess at the court ball.
AL7	This time the crown is seeking the princess in a foreign land.
NH7	This time the princess is wearing the crown at the court ball.
NL7	This time the princess is seeking the crown in a foreign land.
AH8	Today the ship is steering the captain around the dam.
AL8	Today the ship is wiping the captain down for arrival in the port.
NH8	Today the captain is steering the ship around the dam.
NL8	Today the captain is wiping the ship down for arrival in the port.
AH9	Suddenly the safe is robbing the burglar with two old cronies.
AL9	Suddenly the safe is kissing the burglar after finally opening it.
NH9	Suddenly the burglar is robbing the safe with two old cronies.
NL9	Suddenly the burglar is kissing the safe after finally opening it.
AH10	This evening the case is solving the inspector after three months of hard work.
AL10	This evening the case is reading the inspector while the telephone rings.
NH10	This evening the inspector is solving the case after three months of hard work.
NL10	This evening the inspector is reading the case while the telephone rings.

AH11 On a sunny afternoon the flower is picking the girl for the dining table.
AL11 On a sunny afternoon the flower is drawing the girl on a little sketchpad.
NH11 On a sunny afternoon the girl is picking the flower for the dining table.
NL11 On a sunny afternoon the girl is drawing the flower on a little sketchpad.

AH12 Today the mountain is climbing the hiker for the great view.
AL12 Today the mountain is polluting the hiker with the empty packages.
NH12 Today the hiker is climbing the mountain for the great view.
NL12 Today the hiker is polluting the mountain with the empty packages.

AH13 This Thursday the hair is styling the barber with the new hair spray.
AL13 This Thursday the hair is praising the barber with great enthusiasm.
NH13 This Thursday the barber is styling the hair with the new hair spray.
NL13 This Thursday the barber is praising the hair with great enthusiasm.

AH14 In the evening the steak is eating the man on a big plastic plate.
AL14 In the evening the steak is trying the man and the potato salad.
NH14 In the evening the man is eating the steak on a big plastic plate.
NL14 In the evening the man is trying the steak and the potato salad.

AH15 After midnight the movie is watching the boy while lightning strikes.
AL15 After midnight the movie is leaving the boy because of the violent images.
NH15 After midnight the boy is watching the movie while lightning strikes.
AL15 After midnight the boy is leaving the movie because of the violent images.

AH16 In the office the button is pressing the engineer to get help.
AL16 In the office the button is checking the engineer after the horrible accident.
NH16 In the office the engineer is pressing the button to get help.
NL16 In the office the engineer is checking the button after the horrible accident.

AH17 Right now the prayer is saying the priest before the bell rings.
AL17 Right now the prayer is yelling the priest because of a deaf worshipper.
NH17 Right now the priest is saying the prayer before the bell rings.
NL17 Right now the priest is yelling the prayer because of a deaf worshipper.

AH18 This morning the sermon is preaching the pastor with great enthusiasm.
AL18 This morning the sermon is mailing the pastor to the church members.
NH18 This morning the pastor is preaching the sermon with great enthusiasm.
NL18 This morning the pastor is mailing the sermon to the church members.

AH19 Today the rent is paying the lodger for the next two months.
AL19 Today the rent is saving the lodger for the next two months.
NH19 Today the lodger is paying the rent for the next two months.
NL19 Today the lodger is saving the rent for the next two months.

AH20 Before dinner the field is plowing the farmer for the next sowing.
AL20 Before dinner the field is touring the farmer to the check the plants.
NH20 Before dinner the farmer is plowing the field for the next sowing.
NL20 Before dinner the farmer is touring the field to check the plants.

AH21 Friday night the beer is drinking the dude with some good friends.
AL21 Friday night the beer is hiding the dude as the policeman enters.
NH21 Friday night the dude is drinking the beer with some good friends.
NL21 Friday night the dude is hiding the beer as the policeman enters.

AH22 On Saturday the cocktail is mixing the bartender for five pretty girls.
AL22 On Saturday the cocktail is explaining the bartender to a new staff member.
NH22 On Saturday the bartender is mixing the cocktail for five pretty girls.
NL22 On Saturday the bartender is explaining the cocktail to a new staff member.

AH23	In the midday heat the lawn is mowing the gardener with an old scythe.
AL23	In the midday heat the lawn is sweeping the gardener to clear the leaves.
NH23	In the midday heat the gardener is mowing the lawn with an old scythe.
NL23	In the midday heat the gardener is sweeping the lawn to clear the leaves.
AH24	This afternoon the concert is performing the musician without any notes.
AL24	This afternoon the concert is cancelling the musician for inexplicable reasons.
NH24	This afternoon the musician is performing the concert without any notes.
NL24	This afternoon the musician is cancelling the concert for inexplicable reasons.
AH25	Before school the laundry is sorting the mother for the children.
AL25	Before school the laundry is delivering the mother to the neighbors.
NH25	Before school the mother is sorting the laundry for the children.
NL25	Before school the mother is delivering the laundry to the neighbors.
AH26	After dinner the money is losing the gambler that the crime boss won.
AL26	After dinner the money is sending the gambler to the crime boss.
NH26	After dinner the gambler is losing the money that the crime boss won.
NL26	After dinner the gambler is sending the money to the crime boss.
AH27	In the old factory building the painting is framing the artist for the special show.
AL27	In the old factory building the painting is packing the artist after the sales conversation.
NH27	In the old factory building the artist is framing the painting for the special show.
NL27	In the old factory building the artist is packing the painting after the sales conversation.
AH28	In the park the stick is fetching the dog to get a reward.
AL28	In the park the stick is smelling the dog with its excellent nose.
NH28	In the park the dog is fetching the stick to get a reward.
NL28	In the park the dog is smelling the stick with its excellent nose.
AH29	In front of the building the license is examining the bouncer while an angry guest starts shooting.
AL29	In front of the building the license is destroying the bouncer with a big pair of scissors.
NH29	In front of the building the bouncer is destroying the license with a big pair of scissors.
NL29	In front of the building the bouncer is examining the license while an angry guest starts shooting.
AH30	In the morning the damage is repairing the electrician caused by the wind.
AL30	In the morning the damage is faking the electrician to fool the insurance company.
NH30	In the morning the electrician is repairing the damage caused by the wind.
NL30	In the morning the electrician is faking the damage to fool the insurance company.
AH31	After two days the ransom is demanding the kidnapper from the father of the child.
AL31	After two days the ransom is describing the kidnapper for the father of the child.
NH31	After two days the kidnapper is demanding the ransom from the father of the child.
NL31	After two days the kidnapper is describing the ransom for the father of the child.
AH32	Very early in the morning the cake is frosting the confectioner for the wedding on Sunday.
AL32	Very early in the morning the cake is promoting the confectioner as the best in town.
NH32	Very early in the morning the confectioner is frosting the cake for the wedding on Sunday.
NL32	Very early in the morning the confectioner is promoting the cake as the best in town.
AH33	This Wednesday the soup is tasting the chef for the lunch.
AL33	This Wednesday the soup is buying the chef instead of making it from scratch.
NH33	This Wednesday the chef is tasting the soup for the lunch.
NL33	This Wednesday the chef is buying the soup instead of making it from scratch.
AH34	In the evening the vocabulary is learning the schoolgirl with a new memory strategy.
AL34	In the evening the vocabulary is arranging the schoolgirl into different categories.
NH34	In the evening the schoolgirl is learning the vocabulary with a new memory strategy.
NL34	In the evening the schoolgirl is arranging the vocabulary into different categories.

AH35 On the field the foul is calling the referee although the game will end soon.
AL35 On the field the foul is ignoring the referee after the end of the game.
NH35 On the field the referee is calling the foul although the game will end soon.
NL35 On the field the referee is ignoring the foul after the end of the game.

AH36 After a long discussion the oath is swearing the witness before the hearing starts.
AL36 After a long discussion the oath is mumbling the witness before the hearing starts.
NH36 After a long discussion the witness is swearing the oath before the hearing starts.
NL36 After a long discussion the witness is mumbling the oath before the hearing starts.

b) Garden path sentences

GP1 John borrowed the rake or the shovel turned out to be sufficient.
DIFF Might the rake have been borrowed?
EASY Is there a shovel?

GP2 Ms. Haywood planned a picnic or a barbecue was preferred by the kids.
DIFF Might Ms. Haywood have planned a picnic?
EASY Is there a concert?

GP3 The bill arrived in the Senate or the House delayed it again.
DIFF Might the bill have arrived in the Senate?
EASY Is there a bill?

GP4 Linda bought the Chevrolet or the Buick was forced on her by the dealer.
DIFF Might the dealer have forced the Chevrolet on Linda?
EASY Did something happen with a Ferrari?

GP5 Dr. Wendell will perform the surgery or the procedure will be postponed.
DIFF Will the surgery possibly be postponed?
EASY Is there a doctor?

GP6 Liza will perform her famous song or her dance number will end the show.
DIFF Is it Liza's dance number that's famous?
EASY Is there a movie?

GP7 The boys will use the skis or the sled will make the deliveries.
DIFF Will the skis possibly be used?
EASY Are there boys?

GP8 The batter got a hit or a walk was issued intentionally.
DIFF Did the batter definitely get a walk?
EASY Did a building get something?

GP9 William discovered a cure or a vaccine stopped the disease.
DIFF Might William have discovered a cure?
EASY Is there a disease?

GP10 The natives will keep their land or their villages will be taken over.
DIFF Will the natives definitely keep their land?
EASY Is there a shipwreck?

GP11 The space aliens destroyed the ship or the shield remained effective.
DIFF Was the ship possibly destroyed?
EASY Are there aliens?

GP12 Louise talked to the children or the parents scolded them.
DIFF Did Louise talk to the parents?
EASY Did Louise talk to the president?

GP13 The child put on the jacket or the sweater turned out to be warm enough.
DIFF Did the child wear something warm?
EASY Was there a child?

GP14 The factory emitted the waste sludge or the chemicals damaged the river.
DIFF Were the chemicals safe?
EASY Was there a party?

GP15 The tourists visited the shrine or the monastery kept their attention.
DIFF Did they possibly visit the shrine?
EASY Were there tourists?

GP16 The gambler visited the casino or the racetrack proved too much fun.
DIFF Did the gambler dislike the racetrack?
EASY Did the detective do something?

GP17 The team took the train or the subway turned out to be a better option.
DIFF Did the team take public transit?
EASY Did the team do something?

GP18 The lawyer will play the recording or the videotape will sway the jury.
DIFF Will the lawyer film the trial?
EASY Will a concert be recorderd?

GP19 Rachel wrote a poem or an essay filled the space in the magazine.
DIFF Might Rachel write a poem?
EASY Is the girl named Rachel?

GP20 The ants ate through the door or the window was left open.
DIFF Might the ants eat through the window?
EASY Did something happen with apples?

GP21 The animals slept on the straw or the leaves were more comfortable.
DIFF Might the leaves be okay?
EASY Did something happen with animals?

GP22 Richard bought the necklace or the bracelet was a better deal.
DIFF Was the bracelet definitely bought?
EASY Did something happen with a ring?

GP23 The farmers fixed the barn or the outhouse was more important.
DIFF Was the barn possibly fixed?
EASY Did something happen with farmers?

GP24 My dog will chase the ball or the frisbee will keep his attention.
DIFF Might the dog lie down?
EASY Will a cat do something?

c) Relative clause sentences

Object relatives

OR1 The executives that the lawyers sued roused themselves from slumber.
DIFF Was it the executives who roused themselves?
EASY Did a policeman do something?

OR2 The bus driver that the kids followed sifted through the permission slips.
DIFF Did the bus driver follow the kids?
EASY Did a teacher do something?

OR3	The children that the babysitter ignored ranted about their unfair treatment.	
	DIFF	Was it the babysitter that went on a rant?
	EASY	Did a grandmother do something?
OR4	The students that the teacher threatened soiled their pants after the incident.	
	DIFF	Did the students threaten the teacher?
	EASY	Did a director do something?
OR5	The soldiers that the natives helped scaled the big rock that blocked the path.	
	DIFF	Was it the soldiers that scaled the rock?
	EASY	Did a general do something?
OR6	The actor that the director watched soaked the waiter with his drink by accident.	
	DIFF	Did the actor watch the director?
	EASY	Did a chef do something?
OR7	The employees that the fireman spotted berated the EMTs for arriving too late.	
	DIFF	Was it the fireman that berated the EMTs?
	EASY	Did an electrician do something?
OR8	The dancer that the audience loved hustled to the rehearsal that was underway.	
	DIFF	Did the dancer love the audience?
	EASY	Did a musician do something?
OR9	The speaker that the economists entertained revived the crowed after a rousing speech.	
	DIFF	Was it the economists that revived the crowd?
	EASY	Did a reporter do something?
OR10	The ballerina that the murderer saved bestowed many thanks upon her savior.	
	DIFF	Did the ballerina save the murderer?
	EASY	Did a dancer do something?
OR11	The farmer that the rancher hired inquired about new farm equipment.	
	DIFF	Did the rancher hire the farmer?
	EASY	Did a gardener do something?
OR12	The woman that the couple spotted persisted with her one-woman strike.	
	DIFF	Was it the woman who persisted with a strike?
	EASY	Did a boy do something?
OR13	The investigator that the agency phoned vindicated the new person in the accounting department.	
	DIFF	Did the agency phone the investigator?
	EASY	Did a secretary do something?
OR14	The boy that the girl met suppressed his true feelings.	
	DIFF	Was it the girl who suppressed true feelings?
	EASY	Did a woman do something?
OR15	The doctor that the patient saw vanquished all doubt from his mind.	
	DIFF	Did the patient see the doctor?
	EASY	Did a nurse do something?
OR16	The graduate student that the researcher helped relished his time in the lab.	
	DIFF	Was it the graduate student who relished the time?
	EASY	Did a professor do something?
OR17	The man that the woman kissed sauntered confidently into the office.	
	DIFF	Did the woman kiss the man?
	EASY	Did a dog do something?

OR18 The woman that the children irritated loitered around the parking lot.
DIFF Was it the woman who loitered around the parking lot?
EASY Did a man do something?

OR19 The truck driver that the cop accosted pestered law enforcement every chance he got.
DIFF Did the cop accost the truck driver?
EASY Did a porter do something?

OR20 The fisherman that the captain despised recounted all the times he nearly drowned.
DIFF Did the captain despise the fisherman?
EASY Did the fisherman do something?

Subject relatives

SR1 The chef that distracted the waiter sifted the flour onto the counter.
DIFF Did the waiter distract the chef?
EASY Did a chef do something?

SR2 The visitor that introduced the student ranted about rising tuition costs.
DIFF Was it the visitor that went on a rant?
EASY Did a visitor do something?

SR3 The monkeys that watched the zookeepers soiled the bottom of their cage.
DIFF Did the zookeepers watch the monkeys?
EASY Did the monkeys do something?

SR4 The bartender that fought the drunkard scaled the ladder to the roof.
DIFF Was it the drunkard that scaled the ladder?
EASY Did the bartender do something?

SR5 The fireman that signaled the residents soaked the house with high-powered hoses.
DIFF Did the residents signal the fireman?
EASY Did a fireman do something?

SR6 The judge that addressed the witnesses berated the defense attorneys.
DIFF Was it the judge that berated the defense?
EASY Did a judge do something?

SR7 The mathematician that admired the chairman hustled to class across the quad.
DIFF Did the chairman admire the mathematician?
EASY Did a mathematician do something?

SR8 The trainer that called the jockey revived the ailing horse.
DIFF Was it the trainer that revived the horse?
EASY Did a trainer do something?

SR9 The movie star that invited the philanthropists bestowed an annual prize.
DIFF Did the philanthropists invite the movie star?
EASY Did a movie star do something?

SR10 The celebrity that harassed the waiter inquired about getting a bodyguard.
DIFF Was it the waiter who inquired about something?
EASY Did a celebrity do something?

SR11 The professor that ignored the students persisted with his lecture.
DIFF Did the professor ignore the students?
EASY Did a professor do something?

SR12 The manager that visited the foreman vindicated himself at the trial.
DIFF Was it the foreman who vindicated himself?
EASY Did a manager do something?

SR13 The lady that filmed the photographer suppressed all her embarrassing photos.
DIFF Did the lady film the photographer?
EASY Did a lady do something?

SR14 The tennis player that admired the coach vanquished his greatest rival.
DIFF Was it the tennis player who vanquished the rival?
EASY Did a tennis player do something?

SR15 The pilot that delayed the ground crew relished long flight delays.
DIFF Did the pilot delay the ground crew?
EASY Did the pilot do something?

SR16 The wrestler that the boxer challenged sauntered into the ring.
DIFF Was it the boxer who sauntered into the ring?
EASY Did a wrestler do something?

SR17 The mother that the child stopped on the sidewalk loitered idly on the sidewalk.
DIFF Did the mother stop the child?
EASY Did a mother do something?

SR18 The doctor that disliked the patients pestered everyone with his jokes.
DIFF Was it the patients who pestered everyone with jokes?
EASY Did the doctor do something?

SR19 The farmer that harassed the businessman recounts funny stories from his youth all the time.
DIFF Did the farmer harass the businessman?
EASY Did the farmer do something?

2. Sentence material used in Experiment 2

a) Semantic Reversal anomalies

N = non-anomalous, A = anomalous, H = highly associated, L = low associated

NH1 Der Pathologe obduziert den Leichnam nach dem grausamen Verbrechen ganz genau.
NL1 Der Pathologe kühlt den Leichnam nach dem grausamen Verbrechen in der Kältekammer.
AH1 Der Leichnam obduziert den Pathologen nach dem grausamen Verbrechen ganz genau.
AL1 Der Leichnam kühlt den Pathologen nach dem grausamen Verbrechen in der Kältekammer.

NH2 Der Gast trinkt den Wein im Restaurant direkt nach der delikaten Vorspeise.
NL2 Der Gast preist den Wein im Restaurant direkt nach der delikaten Vorspeise.
AH2 Der Wein trinkt den Gast im Restaurant direkt nach der delikaten Vorspeise.
AL2 Der Wein preist den Gast im Restaurant direkt nach der delikaten Vorspeise.

NH3 Der Spieler wirft den Ball eindrucksvoll vor aller Augen in den Basketballkorb.
NL3 Der Spieler reinigt den Ball eindrucksvoll vor aller Augen mit einem pinken Handtuch.
AH3 Der Ball wirft den Spieler eindrucksvoll vor aller Augen in den Basketballkorb.
AL3 Der Ball reinigt den Spieler eindrucksvoll vor aller Augen mit einem pinken Handtuch.

NH4 Der Mechaniker wechselt den Reifen wie jedes Jahr in der Vertragswerkstatt.
NL4 Der Mechaniker leiht den Reifen wie jedes Jahr in der Vertragswerkstatt.
AH4 Der Reifen wechselt den Mechaniker wie jedes Jahr in der Vertragswerkstatt.
AL4 Der Reifen leiht den Mechaniker wie jedes Jahr in der Vertragswerkstatt.

NH5 Der Pendler kauft den Fahrschein unmittelbar vor Fahrtantritt am Bahnschalter.
NL5 Der Pendler tauscht den Fahrschein unmittelbar vor Fahrtantritt am Bahnschalter.
AH5 Der Fahrschein kauft den Pendler unmittelbar vor Fahrtantritt am Bahnschalter.
AL5 Der Fahrschein tauscht den Pendler unmittelbar vor Fahrtantritt am Bahnschalter.

NH6 Der Rocker raucht den Joint verbotenerweise auf der Parkhaustoilette.
NL6 Der Rocker kauft den Joint verbotenerweise auf der Parkhaustoilette.
AH6 Der Joint raucht den Rocker verbotenerweise auf der Parkhaustoilette.
AL6 Der Joint kauft den Rocker verbotenerweise auf der Parkhaustoilette.

NH7 Der Stürmer schießt den Elfmeter nach dem fiesen Foul direkt in die linke Torecke.
NL7 Der Stürmer fordert den Elfmeter nach dem fiesen Foul als Entschädigung.
AH7 Der Elfmeter schießt den Stürmer nach dem fiesen Foul direkt in die linke Torecke.
AL7 Der Elfmeter fordert den Stürmer nach dem fiesen Foul als Entschädigung.

NH8 Der Page packt den Koffer für die Reise des Präsidenten nach Dubai.
NL8 Der Page bemalt den Koffer für die Reise des Präsidenten nach Dubai.
AH8 Der Koffer packt den Pagen für die Reise des Präsidenten nach Dubai.
AL8 Der Koffer bemalt den Pagen für die Reise des Präsidenten nach Dubai.

NH9 Der Handwerker verlegt den Fußboden wegen des nahenden Einzugstermins an einem einzigen Nachmittag.
NL9 Der Handwerker zimmert den Fußboden wegen des nahenden Einzugstermins aus alten Sperrholzplatten.
AH9 Der Fußboden verlegt den Handwerker wegen des nahenden Einzugstermins an einem einzigen Nachmittag.
AL9 Der Fußboden zimmert den Handwerker wegen des nahenden Einzugstermins aus alten Sperrholzplatten.

NH10 Der Lektor liest den Roman bereits vor der öffentlichen Vorstellung auf der Buchmesse.
NL10 Der Lektor bezahlt den Roman bereits vor der öffentlichen Vorstellung auf der Buchmesse.
AH10 Der Roman liest den Lektor bereits vor der öffentlichen Vorstellung auf der Buchmesse.
AL10 Der Roman bezahlt den Lektor bereits vor der öffentlichen Vorstellung auf der Buchmesse.

NH11 Der Pirat findet den Schatz vor der Küste Madagaskars in einer schwer zugänglichen Bucht.
NL11 Der Pirat angelt den Schatz vor der Küste Madagaskars in einer schwer zugänglichen Bucht.
AH11 Der Schatz findet den Piraten vor der Küste Madagaskars in einer schwer zugänglichen Bucht.
AL11 Der Schatz angelt den Piraten vor der Küste Madagaskars in einer schwer zugänglichen Bucht.

NH12	Der Schüler schreibt den Aufsatz in einem schlechten und kaum verständlichen Deutsch.
NL12	Der Schüler klaut den Aufsatz leider in einem schlechten Augenblick, nämlich als der Direktor den Raum betritt.
AH12	Der Aufsatz schreibt den Schüler leider in einem schlechten und kaum verständlichen Deutsch.
AL12	Der Aufsatz klaut den Schüler leider in einem schlechten Augenblick, nämlich als der Direktor den Raum betritt.
NH13	Der Gärtner verursacht den Schaden dummerweise bei einer nächtlichen Rauschfahrt mit dem Rasentraktor.
NL13	Der Gärtner leugnet den Schaden dummerweise bei einem Verhör auf der örtlichen Polizeiwache.
AH13	Der Schaden verursacht den Gärtner dummerweise bei einer nächtlichen Rauschfahrt mit dem Rasentraktor.
AL13	Der Schaden leugnet den Gärtner dummerweise bei einem Verhör auf der örtlichen Polizeiwache.
NH14	Der Dieb knackt den Tresor bei einem Überfall mit einem Spezialdetektor aus NSA-Beständen.
NL14	Der Dieb besprüht den Tresor bei einem Überfall mit hochexplosivem Spezialspray.
AH14	Der Tresor knackt den Dieb bei einem Überfall mit einem Spezialdetektor aus NSA-Beständen.
AL14	Der Tresor besprüht den Dieb bei einem Überfall mit hochexplosivem Spezialspray.
NH15	Der Azubi kocht den Kaffee als morgendliches Ritual für die gesamte Personalabteilung.
NL15	Der Azubi trägt den Kaffee als morgendliches Ritual in das Büro des Geschäftsführers.
AH15	Der Kaffee kocht den Azubi als morgendliches Ritual für die gesamte Personalabteilung.
AL15	Der Kaffee trägt den Azubi als morgendliches Ritual in das Büro des Geschäftsführers.
NH16	Der Pfleger schiebt den Rollstuhl nun schon zum zweiten Mal zur Intensivstation.
NL16	Der Pfleger montiert den Rollstuhl nun schon zum zweiten Mal mit dem kaputten Schraubenzieher.
AH16	Der Rollstuhl schiebt den Pfleger nun schon zum zweiten Mal zur Intensivstation.
AL16	Der Rollstuhl montiert den Pfleger nun schon zum zweiten Mal mit dem kaputten Schraubenzieher.
NH17	Der Vater zahlt den Unterhalt nach einem Gerichtsbeschluss bis zum Ende der Ausbildung.
NL17	Der Vater spart den Unterhalt nach einem Gerichtsbeschluss für die Finanzierung der Ausbildung.
AH17	Der Unterhalt zahlt den Vater nach einem Gerichtsbeschluss bis zum Ende der Ausbildung.
AL17	Der Unterhalt spart den Vater nach einem Gerichtsbeschluss für die Finanzierung der Ausbildung.
NH18	Der Redner hält den Vortrag auf dem Kongress erst nach dem Abendessen.
NL18	Der Redner schreibt den Vortrag auf dem Kongress erst nach dem Abendessen.
AH18	Der Vortrag hält den Redner auf dem Kongress erst nach dem Abendessen.
AL18	Der Vortrag schreibt den Redner auf dem Kongress erst nach dem Abendessen.
NH19	Der Läufer bricht den Rekord überraschenderweise schon beim Vorlauf über die 100 Meter.
NL19	Der Läufer vergisst den Rekord überraschenderweise beim Interview mit der Reporterin.
AH19	Der Rekord bricht den Läufer überraschenderweise schon beim Vorlauf über die 100 Meter.
AL19	Der Rekord vergisst den Läufer überraschenderweise beim Interview mit der Reporterin.
NH20	Der Opa bucht den Flug in diesem Jahr zum ersten Mal über ein Onlineportal.
NL20	Der Opa fürchtet den Flug in diesem Jahr ganz besonders wegen des angesagten Schneefalls.
AH20	Der Flug bucht den Opa in diesem Jahr zum ersten Mal über ein Onlineportal.
AL20	Der Flug fürchtet den Opa in diesem Jahr ganz besonders wegen des angesagten Schneefalls.
NH21	Der Sieger verteidigt den Titel auch bei diesem Rennen mit großer Souveränität.
NL21	Der Sieger widmet den Titel auch bei diesem Rennen einem verstorbenen Freund aus Schulzeiten.
AH21	Der Titel verteidigt den Sieger auch bei diesem Rennen mit großer Souveränität.
AL21	Der Titel widmet den Sieger auch bei diesem Rennen einem verstorbenen Freund aus Schulzeiten.
NH22	Der Mann rasiert den Bart heute mit dem neuen Rasierapparat aus der Werbung.
NL22	Der Mann färbt den Bart heute mit dem neuen Haarfärbemittel aus der Werbung.
AH22	Der Bart rasiert den Mann heute mit dem neuen Rasierapparat aus der Werbung.
AL22	Der Bart färbt den Mann heute mit dem neuen Haarfärbemittel aus der Werbung.
NH23	Der Junge spitzt den Bleistift aus Frust über die Lehrerin mit der Bastelschere.
NL23	Der Junge zerschneidet den Bleistift aus Frust über die Lehrerin mit der Bastelschere.
AH23	Der Bleistift spitzt den Jungen aus Frust über die Lehrerin mit der Bastelschere.
AL23	Der Bleistift zerschneidet den Jungen aus Frust über die Lehrerin mit der Bastelschere.

NH24 Der Sportler bindet den Schuh mit einem kniffligen Knoten aus der letzten Lektion des Segelkurses.
NL24 Der Sportler holt den Schuh mit einem kniffligen Trick aus dem obersten Regal.
AH24 Der Schuh bindet den Sportler mit einem kniffligen Knoten aus der letzten Lektion des Segelkurses.
AL24 Der Schuh holt den Sportler mit einem kniffligen Trick aus dem obersten Regal.

NH25 Der Besucher hört den Klang bereits beim Betreten der alten, majestätisch aufragenden Domkapelle.
NL25 Der Besucher erahnt den Klang bereits beim Betreten der alten, majestätisch aufragenden Domkapelle.
AH25 Der Klang hört den Besucher bereits beim Betreten der alten, majestätisch aufragenden Domkapelle.
AL25 Der Klang erahnt den Besucher bereits beim Betreten der alten, majestätisch aufragenden Domkapelle.

NH26 Der Dieb gräbt den Tunnel nach monatelanger Planung durch den Keller des Nachbarhauses hindurch.
NL26 Der Dieb findet den Tunnel nach monatelanger Planung im Keller des Nachbarhauses.
AH26 Der Tunnel gräbt den Dieb nach monatelanger Planung durch den Keller des Nachbarhauses hindurch.
AL26 Der Tunnel findet den Dieb nach monatelanger Planung im Keller des Nachbarhauses.

NH27 Der Dozent betritt den Raum mit dem Klausurenstapel in der Hand.
NL27 Der Dozent versperrt den Raum mit dem Klausurenstapel in der Hand.
AH27 Der Raum betritt den Dozenten mit dem Klausurenstapel in der Hand.
AL27 Der Raum versperrt den Dozenten mit dem Klausurenstapel in der Türe.

NH28 Der Forscher erbringt den Nachweis schließlich doch noch vor dem Auslaufen des Drittmittelprojektes.
NL28 Der Forscher malt den Nachweis schließlich doch in verkürzter Form auf die Tafel.
AH28 Der Nachweis erbringt den Forscher schließlich doch noch vor dem Auslaufen des Drittmittelprojektes.
AL28 Der Nachweis malt den Forscher schließlich doch in verkürzter Form auf die Tafel.

NH29 Der Bäcker knetet den Teig mehrmals für einige Minuten kräftig durch, damit der Kuchen später locker wird.
NL29 Der Bäcker erwärmt den Teig mehrmals für einige Minuten, damit der Kuchen später gut aufgeht.
AH29 Der Teig knetet den Bäcker mehrmals für einige Minuten kräftig durch, damit der Kuchen später locker wird.
AL29 Der Teig erwärmt den Bäcker mehrmals für einige Minuten, damit der Kuchen später gut aufgeht.

NH30 Der Mentor gibt den Ratschlag erst nach einiger Überwindung dem verängstigten Prüfungskandidaten.
NL30 Der Mentor liest den Ratschlag erst nach einiger Überwindung in einem alten Selbsthilfebuch.
AH30 Der Ratschlag gibt den Mentor erst nach einiger Überwindung dem verängstigten Prüfungskandidaten.
AL30 Der Ratschlag liest den Mentor erst nach einiger Überwindung in einem alten Selbsthilfebuch.

NH31 Der Therapeut löst den Konflikt manchmal nur durch aufmerksames Zuhören und vorsichtiges Vermitteln.
NL31 Der Therapeut kennt den Konflikt manchmal nur durch aufmerksames Zuhören bei anderen Gesprächen.
AH31 Der Konflikt löst den Therapeuten manchmal nur durch aufmerksames Zuhören und vorsichtiges Vermitteln.
AL31 Der Konflikt kennt den Therapeuten manchmal nur durch aufmerksames Zuhören bei anderen Gesprächen.

NH32 Der Redakteur verfasst den Beitrag nach dem Willen des Herausgebers über die Bienenzucht.
NL32 Der Redakteur kopiert den Beitrag nach dem Willen des Herausgebers aus einer alten Ausgabe.
AH32 Der Beitrag verfasst den Redakteur nach dem Willen des Herausgebers über die Bienenzucht.
AL32 Der Beitrag kopiert den Redakteur nach dem Willen des Herausgebers aus einer alten Ausgabe.

NH33 Der Landwirt erntet den Spargel wegen des anhaltend guten Wetters schon Anfang April.
NL33 Der Landwirt bringt den Spargel wegen des anhaltend guten Wetters auf die Terrasse.
AH33 Der Spargel erntet den Landwirt wegen des anhaltend guten Wetters schon Anfang April.
AL33 Der Spargel bringt den Landwirt wegen des anhaltend guten Wetters auf die Terrasse.

NH34 Der Kunde äußert den Wunsch zunächst nur zaghaft, dann aber mit Nachdruck.
NL34 Der Kunde verschweigt den Wunsch zunächst nur ungern, dann aber scheint es die bessere Alternative.
AH34 Der Wunsch äußert den Kunden zunächst nur zaghaft, dann aber mit Nachdruck.
AL34 Der Wunsch verschweigt den Kunden zunächst nur ungern, dann aber scheint es die bessere Alternative.

NH35 Der Kritiker schaut den Film bei der Premiere mit einer gehörigen Portion Skepsis.
NL35 Der Kritiker mag den Film bei der Premiere nicht besonders, vor allem wegen der schlechten Regieführung.
AH35 Der Film schaut den Kritiker bei der Premiere mit einer gehörigen Portion Skepsis.
AL35 Der Film mag den Kritiker bei der Premiere nicht besonders, vor allem wegen der schlechten Regieführung.

NH36 Der Lehrling schält den Apfel manchmal so langsam, dass die Köchin ganz nervös wird.
NL36 Der Lehrling putzt den Apfel manchmal so langsam, dass die Köchin ganz nervös wird.
AH36 Der Apfel schält den Lehrling manchmal so langsam, dass die Köchin ganz nervös wird.
AL36 Der Apfel putzt den Lehrling manchmal so langsam, dass die Köchin ganz nervös wird.

NH37 Der Ingenieur baut den Turm am östlichen Stadtzentrum genau nach den Entwürfen des Architekten.
NL37 Der Ingenieur sieht den Turm am östlichen Stadtzentrum kurz nach Abschluss der dreijährigen Bauphase.
AH37 Der Turm baut den Ingenieur am östlichen Stadtzentrum genau nach den Entwürfen des Architekten.
AL37 Der Turm sieht den Ingenieur am östlichen Stadtzentrum kurz nach Abschluss der dreijährigen Bauphase.

NH38 Der Heizer fegt den Schornstein natürlich nicht aus purem Vergnügen.
NL38 Der Heizer saugt den Schornstein natürlich nicht aus purem Vergnügen.
AH38 Der Schornstein fegt den Heizer natürlich nicht aus purem Vergnügen.
AL38 Der Schornstein saugt den Heizer natürlich nicht aus purem Vergnügen.

NH39 Der Platzwart mäht den Rasen wegen des morgigen Fußballspiels schon zum zweiten Mal in dieser Woche.
NL39 Der Platzwart schützt den Rasen wegen des morgigen Fußballspiels mit einer Spezialplane.
AH39 Der Rasen mäht den Platzwart wegen des morgigen Fußballspiels schon zum zweiten Mal in dieser Woche.
AL39 Der Rasen schützt den Platzwart wegen des morgigen Fußballspiels mit einer Spezialplane.

NH40 Der Arbeiter rodet den Wald auf Anordnung des Forstamtes, um neue Bauflächen zu erschließen.
NL40 Der Arbeiter pflanzt den Wald auf Anordnung des Forstamtes, um ein Naherholungsgebiet zu schaffen.
AH40 Der Wald rodet den Arbeiter auf Anordnung des Forstamtes, um neue Bauflächen zu erschließen.
AL40 Der Wald pflanzt den Arbeiter auf Anordnung des Forstamtes, um ein Naherholungsgebiet zu schaffen.

NH41 Der Student isst den Döner in der Oberstadt immer mit Knoblauchsauce.
NL41 Der Student liebt den Döner in der Oberstadt besonders mit Knoblauchsauce.
AH41 Der Döner isst den Studenten in der Oberstadt immer mit Knoblauchsauce.
AL41 Der Döner liebt den Studenten in der Oberstadt besonders mit Knoblauchsauce.

NH42 Der Verwalter bekämpft den Schimmelpilz nach einem Hinweis der Mieterin mit einem chemischen Alleskönner.
NL42 Der Verwalter bemerkt den Schimmelpilz nach einem Hinweis der Mieterin dann auch selbst.
AH42 Der Schimmelpilz bekämpft den Verwalter nach einem Hinweis der Mieterin mit einem chemischen Alleskönner.
AL42 Der Schimmelpilz bemerkt den Verwalter nach einem Hinweis der Mieterin dann auch selbst.

NH43 Der Extremist verübt den Terroranschlag mit einer bis zu diesem Zeitpunkt unbekannten Grausamkeit.
NL43 Der Extremist verteidigt den Terroranschlag mit einer fanatischen Botschaft auf Youtube.
AH43 Der Terroranschlag verübt den Extremisten mit einer bis zu diesem Zeitpunkt unbekannten Grausamkeit.
AL43 Der Terroranschlag verteidigt den Extremisten mit einer fanatischen Botschaft auf Youtube.

NH44 Der Komiker erzählt den Witz in der Fernsehsendung mit beeindruckender Mimik.
NL44 Der Komiker würdigt den Witz in der Fernsehsendung mit einem Augenzwinkern.
AH44 Der Witz erzählt den Komiker in der Fernsehsendung mit beeindruckender Mimik.
AL44 Der Witz würdigt den Komiker in der Fernsehsendung mit einem Augenzwinkern.

NH45 Der Clown verliert den Überblick im ganzen Wirrwarr von Veranstaltungspunkten.
NL45 Der Clown präsentiert den Überblick im ganzen Wirrwarr von Veranstaltungspunkten.
AH46 Der Überblick verliert den Clown im ganzen Wirrwarr von Veranstaltungspunkten.
AL46 Der Überblick präsentiert den Clown im ganzen Wirrwarr von Veranstaltungspunkten.

NH46 Der Feuerwehrmann löscht den Brand manchmal gar nicht mit Wasser, sondern nur mit Schaum.
NL46 Der Feuerwehrmann riecht den Brand manchmal gar nicht, sondern hört nur ein Knistern im Gebälk.
AH46 Der Brand löscht den Feuerwehrmann manchmal gar nicht mit Wasser, sondern nur mit Schaum.
AH46 Der Brand riecht den Feuerwehrmann manchmal gar nicht, sondern hört nur ein Knistern im Gebälk.

NH47 Der Pförtner entsorgt den Müll im Abfallcontainer neben der Hofeinfahrt.
NL47 Der Pförtner entflammt den Müll im Abfallcontainer neben der Hofeinfahrt.
AH47 Der Müll entsorgt den Pförtner im Abfallcontainer neben der Hofeinfahrt.
AL47 Der Müll entflammt den Pförtner im Abfallcontainer neben der Hofeinfahrt.

NH48 Der Tourist verlängert den Aufenthalt noch um zwei Wochen, weil das Wetter so toll ist.
NL48 Der Tourist verschiebt den Aufenthalt noch um zwei Wochen, weil das Wetter so toll ist.
AH48 Der Aufenthalt verlängert den Touristen noch um zwei Wochen, weil das Wetter so toll ist.
AL48 Der Aufenthalt verschiebt den Touristen noch um zwei Wochen, weil das Wetter so toll ist.

NH49 Der Arzt zieht den Zahn nach der Wurzelbehandlung ohne Vorwarnung heraus.
NL49 Der Arzt poliert den Zahn nach der Wurzelbehandlung noch einmal ganz gründlich.
AH49 Der Zahn zieht den Arzt nach der Wurzelbehandlung ohne Vorwarnung heraus.
AL49 Der Zahn poliert den Arzt nach der Wurzelbehandlung noch einmal ganz gründlich.

NH50 Der Chirurg amputiert den Arm bei einem Knochentumor erst als letzte Option.
NL50 Der Chirurg bestreicht den Arm bei einem Knochentumor mit einer radioaktiven Paste.
AH50 Der Arm amputiert den Chirurgen bei einem Knochentumor erst als letzte Option.
AL50 Der Arm bestreicht den Chirurgen bei einem Knochentumor mit einer radioaktiven Paste.

NH51 Der Verehrer pflückt den Blumenstrauß immer im Blumenbeet vor dem Stadtbüro.
NL51 Der Verehrer stiehlt den Blumenstrauß immer im Blumenbeet vor dem Stadtbüro.
AH51 Der Blumenstrauß pflückt den Verehrer immer im Blumenbeet vor dem Stadtbüro.
AL51 Der Blumenstrauß stiehlt den Verehrer immer im Blumenbeet vor dem Stadtbüro.

NH52 Der Weltmeister feiert den Sieg trotz des schweren Unfalls im Finale ausgelassen bis zum frühen Morgen.
NL52 Der Weltmeister erwartet den Sieg trotz des schweren Unfalls im letzten Trainingslager.
AH52 Der Sieg feiert den Weltmeister trotz des schweren Unfalls im Finale ausgelassen bis zum frühen Morgen.
AL52 Der Sieg erwartet den Weltmeister trotz des schweren Unfalls im letzten Trainingslager.

NH53 Der Flüchtling beantragt den Pass bei der Ausländerbehörde, um dauerhaft in Deutschland bleiben zu können.
NL53 Der Flüchtling fordert den Pass bei der Ausländerbehörde, um dauerhaft in Deutschland bleiben zu können.
AH53 Der Pass beantragt den Flüchtling bei der Ausländerbehörde, um dauerhaft in Deutschland bleiben zu können.
AL53 Der Pass fordert den Flüchtling bei der Ausländerbehörde, um dauerhaft in Deutschland bleiben zu können.

NH54 Der General führt den Krieg auch aus persönlichen Motiven heraus bis zum bitteren Ende.
NL54 Der General billigt den Krieg auch aus persönlichen Motiven heraus.
AH54 Der Krieg führt den General auch aus persönlichen Motiven heraus bis zum bitteren Ende.
AL54 Der Krieg billigt den General auch aus persönlichen Motiven heraus.

NH55 Der Beamte leert den Papierkorb pflichtbewusst jeden Freitagnachmittag.
NL55 Der Beamte füllt den Papierkorb pflichtbewusst jeden Freitagnachmittag mit alten Unterlagen.
AH55 Der Papierkorb leert den Beamten pflichtbewusst jeden Freitagnachmittag.
AL55 Der Papierkorb füllt den Beamten pflichtbewusst jeden Freitagnachmittag mit alten Unterlagen.

NH56 Der Bauer pflügt den Acker mehrmals, um ein gutes Anwachsen der Saat zu gewährleisten.
NL56 Der Bauer kontrolliert den Acker mehrmals, um ein gutes Anwachsen der Saat zu gewährleisten.
AH56 Der Acker pflügt den Bauern mehrmals, um ein gutes Anwachsen der Saat zu gewährleisten.
AL56 Der Acker kontrolliert den Bauern mehrmals, um ein gutes Anwachsen der Saat zu gewährleisten.

NH57 Der Onkel pflegt den Kontakt nur sporadisch, auch aufgrund der Entfernung.
NL57 Der Onkel erträgt den Kontakt nur sporadisch, beispielsweise zu den wichtigen Feiertagen.
AH57 Der Kontakt pflegt den Onkel nur sporadisch, auch aufgrund der Entfernung.
AL57 Der Kontakt erträgt den Onkel nur sporadisch, beispielsweise zu den wichtigen Feiertagen.

NH58 Der Soldat verweigert den Gehorsam auch im Gefecht, wenn es darum geht, einen Menschen zu erschießen.
NL58 Der Soldat erklärt den Gehorsam auch im Gefecht, wenn es darum geht, einen Menschen zu erschießen.
AH58 Der Gehorsam verweigert den Soldaten auch im Gefecht, wenn es darum geht, einen Menschen zu erschießen.
AL58 Der Gehorsam erklärt den Soldaten auch im Gefecht, wenn es darum geht, einen Menschen zu erschießen.

NH59 Der Athlet läuft den Marathon trotz einer chronischen Muskelentzündung in der linken Wade.
NL59 Der Athlet übt den Marathon trotz einer chronischen Muskelentzündung in der linken Wade.
AH59 Der Marathon läuft den Athleten trotz einer chronischen Muskelentzündung in der linken Wade.
AL59 Der Marathon übt den Athleten trotz einer chronischen Muskelentzündung in der linken Wade.

NH60	Der Künstler bepflanzt den Garten nach französischer Tradition mit zahlreichen südländischen Sträuchern.
NL60	Der Künstler teilt den Garten nach französischer Tradition in mehrere symmetrische Zirkel.
AH60	Der Garten bepflanzt den Künstler nach französischer Tradition mit zahlreichen südländischen Sträuchern.
AL60	Der Garten teilt den Künstler nach französischer Tradition in mehrere symmetrische Zirkel.
NH61	Der Apotheker lindert den Schmerz durch ein Blütenextrakt der südamerikanischen Graviola-Pflanze.
NL61	Der Apotheker steigert den Schmerz durch ein Blütenextrakt der südamerikanischen Graviola-Pflanze.
AH61	Der Schmerz lindert den Apotheker durch ein Blütenextrakt der südamerikanischen Graviola-Pflanze.
AL61	Der Schmerz steigert den Apotheker durch ein Blütenextrakt der südamerikanischen Graviola-Pflanze.
NH62	Der Wanderer besteigt den Berg kurz vor Sonnenaufgang, um das Lichtspiel zu beobachten.
NL62	Der Wanderer zeichnet den Berg kurz vor Sonnenaufgang, um das Lichtspiel zu dokumentieren.
AH62	Der Berg besteigt den Wanderer kurz vor Sonnenaufgang, um das Lichtspiel zu beobachten.
AL62	Der Berg zeichnet den Wanderer kurz vor Sonnenaufgang, um das Lichtspiel zu dokumentieren.
NH63	Der Oberst erteilt den Befehl erst nach der Rückversicherung beim Heerführer.
NL63	Der Oberst versteht den Befehl erst nach der Rückfrage beim Heerführer.
AH63	Der Befehl erteilt den Oberst erst nach der Rückversicherung beim Heerführer.
AL63	Der Befehl versteht den Oberst erst nach der Rückfrage beim Heerführer.
NH64	Der Koch öffnet den Deckel nur ganz vorsichtig, um den Gahrungsprozess nicht zu gefährden.
NL64	Der Koch spült den Deckel nur ganz vorsichtig, um die Spezialbeschichtung nicht zu zerstören.
AH64	Der Deckel öffnet den Koch nur ganz vorsichtig, um den Gahrungsprozess nicht zu gefährden.
AL64	Der Deckel spült den Koch nur ganz vorsichtig, um die Spezialbeschichtung nicht zu zerstören.
NH65	Der Lehrer korrigiert den Fehler nach zwanzig Berufsjahren schon fast automatisch.
NL65	Der Lehrer errät den Fehler nach zwanzig Berufsjahren schon fast automatisch.
AH65	Der Fehler korrigiert den Lehrer nach zwanzig Berufsjahren schon fast automatisch.
AL65	Der Fehler errät den Lehrer nach zwanzig Berufsjahren schon fast automatisch.
NH66	Der Mieter unterschreibt den Vertrag wegen der Prämie schließlich doch.
NL66	Der Mieter zerreißt den Vertrag wegen der unverschämten Konditionen auf der Stelle.
AH66	Der Vertrag unterschreibt den Mieter wegen der Prämie schließlich doch.
AL66	Der Vertrag erhält den Mieter wegen der unverschämten Konditionen auf der Stelle.
NH67	Der Kapitän lichtet den Anker angesichts des herannahenden Gewitters hastiger als gewöhnlich.
NL67	Der Kapitän sucht den Anker angesichts des herannahenden Gewitters im Chaos der Schiffstaue.
AH67	Der Anker lichtet den Kapitän angesichts des herannahenden Gewitters hastiger als gewöhnlich.
AL67	Der Anker sucht den Kapitän angesichts des herannahenden Gewitters im Chaos der Schiffstaue.
NH68	Der Indianer spannt den Bogen mit dem rechten Arm so weit, dass die Sehne reißt.
NL68	Der Indianer dehnt den Bogen mit dem rechten Arm so weit, dass die Sehne reißt.
AH68	Der Bogen spannt den Indianer mit dem rechten Arm so weit, dass die Sehne reißt.
AL68	Der Bogen dehnt den Indianer mit dem rechten Arm so weit, dass die Sehne reißt.
NH69	Der Maurer trägt den Helm aufgrund der Sicherheitsrichtlinien jetzt auch während der Mittagspause.
NL69	Der Maurer ergreift den Helm aufgrund der Sicherheitsrichtlinien mit beiden Händen.
AH69	Der Helm trägt den Maurer aufgrund der Sicherheitsrichtlinien jetzt auch während der Mittagspause.
AL69	Der Helm ergreift den Maurer aufgrund der Sicherheitsrichtlinien mit beiden Händen.
NH70	Der Ritter stirbt den Heldentod beim Kampf gegen die königlichen Hilfstruppen.
NL70	Der Ritter würdigt den Heldentod beim Kampf gegen die königlichen Hilfstruppen.
AH70	Der Heldentod stirbt den Ritter beim Kampf gegen die königlichen Hilfstruppen.
AL70	Der Ritter stirbt den Heldentod beim Kampf gegen die königlichen Hilfstruppen.
NH71	Der Butler deckt den Tisch heute ausnahmsweise mit dem teuren Meißner Porzellan.
NL71	Der Butler stellt den Tisch heute ausnahmsweise an einen anderen Ort.
AH71	Der Tisch deckt den Butler heute ausnahmsweise mit dem teuren Meißner Porzellan.
AL71	Der Tisch stellt den Butler heute ausnahmsweise an einen anderen Ort.

NH72 Der Monteur zieht den Stecker grundsätzlich nach jedem Benutzen der Stichsäge.
NL72 Der Monteur verklebt den Stecker grundsätzlich nach jedem Benutzen der Stichsäge.
AH72 Der Stecker zieht den Monteur grundsätzlich nach jedem Benutzen der Stichsäge.
AL72 Der Stecker verklebt den Monteur grundsätzlich nach jedem Benutzen der Stichsäge.

b) Filler sentences

Non-anomalous sentences with highly predictable target word of low frequency

NHL1 Als Jonas heimlich die Treppe runterläuft, knarzt eine alte ***Stufe*** unter seinem Gewicht.
NHL2 Im Juweliergeschäft kauft sich die junge Frau ein Paar edle ***Ohrringe*** und ein goldenes Armband.
NHL3 Um das Wasser aus der Badewanne zu lassen, zieht Jan den ***Stöpsel*** aus dem Abfluss.
NHL4 Das Kleid war schneeweiß und unten am ***Saum*** waren rote Rosen eingestickt.
NHL5 Um den Kopf zu schützen, sollten man in der Sonne immer einen ***Hut*** tragen.
NHL6 Im Vorstellungsgespräch trägt die Abiturientin einen dunklen Rock und eine weiße ***Bluse*** mit Rüschen.
NHL7 Das Ehepaar frühstückt jeden Morgen vor dem Haus auf der ***Terrasse*** und genießt das Vögelgezwitscher.
NHL8 Der Friseur empfiehlt, mit den Haaren nach dem Duschen regelmäßig eine ***Kur*** zu machen.
NHL9 Bei der Gitarre reißt beim Spielen manchmal auch eine hochwertige ***Saite*** und schnellt durch die Luft.
NHL10 Weil sie Menschenfleisch aßen, wurden die Ureinwohner nur als ***Kannibalen*** bezeichnet.
NHL11 Die Schiffspassagiere stehen an der ***Reling*** und winken zum Abschied den Freunden und Familien zu.
NHL12 Obwohl Lara nicht daran glaubt, liest sie in der Zeitung immer zuerst ihr ***Horoskop*** für die nächste Woche.
NHL13 Der wohl typischste Backwarenexport aus Bayern ist die ***Brezel***, welche in ganz Deutschland beliebt ist.
NHL14 Nach zwei Tagen ohne Warmwasser wird der defekte ***Boiler*** durch einen neuen ersetzt.
NHL15 Das Schriftsystem der alten Ägypter bestand aus zahlreichen ***Hieroglyphen*** in verschiedenen Größen.
NHL16 Das Wasser des Springbrunnens schießt wie eine ***Fontäne*** in die Luft.
NHL17 Damit wir im Dunkeln sehen können, weitet sich unsere ***Pupille*** und passt sich dem Lichteinfall an.
NHL18 Sven bringt den Biomüll auf den ***Kompost*** im hinteren Teil des Gemüsegartens.
NHL19 Buchstaben werden eingeteilt in Konsonanten und ***Vokale***, wobei man letztere auch als „Selbstlaute“ bezeichnet.
NHL20 Nach dem Regen liebt es Timmy, mit den Gummistiefeln in die großen ***Pfützen*** zu springen.

Non-anomalous sentences with low predictable target word of low frequency

NLL1 Matthias kennt sie genau, die alte ***Stufe*** im Flur, die immer knarzt.
NLL2 Auf dem Flohmarkt um die Ecke kauft Udo wunderschöne ***Ohrringe*** für seine Frau.
NLL3 Während sich neben der Spüle die Teller türmen, kann Jana einfach den ***Stöpsel*** für den Abfluss nicht finden.
NLL4 Schon lange war Anja bei der Arbeit, aber der ***Saum*** des Kleides gefiel ihr immer noch nicht.
NLL5 Das Geschäft gewährt jedem Kunden beim Kauf eines ***Hutes*** zwanzig Prozent Rabatt.
NLL6 Da ihre Lieblingsfarbe rot ist, kann Anja weiße ***Blusen*** absolut nicht leiden.
NLL7 Der Welpe bekam eine solche Angst, dass er sich unter der ***Terrasse*** versteckte.
NLL8 Alle bewundern Sophies Haare, aber das liegt nur daran, dass sie vorher eine ***Kur*** gemacht hat.
NLL9 Es gibt kleine Manufakturen, die darauf spezialisiert sind, hochwertige ***Saiten*** für Konzertgitarren herzustellen.
NLL10 Die Ureinwohner verstehen nicht, warum man sie als ***Kannibalen*** bezeichnet.
NLL11 René verliert das Gleichgewicht und hält sich an der ***Reling*** fest, um nicht über Bord zu gehen.
NLL12 Viele wissen es nicht, aber dass Susi immer ihr ***Horoskop*** liest, ist keineswegs so harmlos, wie es scheint.
NLL13 Beim Einkaufen freut sich Bernd stets auf die ***Brezel***, die seine Mutter ihm für den Heimweg kauft.
NLL14 Die Reparatur des defekten ***Boilers*** beansprucht mehr Zeit als gedacht.
NLL15 Manchmal erschien Klaus das, was er sah, wie zahlreiche ***Hieroglyphen*** in verschiedenen Größen.
NLL16 Der Passant zeigt den Touristen den Weg zu einer ***Fontäne*** vor dem Stadtschloss.
NLL17 Der menschliche Körper ist einfach perfekt konstruiert, schon unsere ***Pupille*** ist ein einziges Wunderwerk.
NLL18 Der Nachbar beschwert sich regelmäßig über den Gestank des ***Komposts*** im Nachbargarten.
NLL19 Selbst Muttersprachlern fällt es schwer, Diphthonge und ***Vokale*** richtig zu unterscheiden.
NLL20 Ihre Mutter wird sofort zornig, als Martina freudig in die großen ***Pfützen*** springt.

Non-anomalous sentences with highly predictable target word of high frequency

NHH1 Um ihren Sohn anzurufen, greift Hilde zu ihrem schnurlosen ***Telefon*** und tippt die Nummer ein.
NHH2 Der Pianist spielt das Lied auf dem ***Klavier*** einmal vor, damit alle es hören.
NHH3 Vor 40 Jahren gaben die Beatles ihr berühmtes ***Konzert*** im ausverkauften Prince-of-Wales-Theatre in London.
NHH4 Wegen der stickigen Luft öffnet Hans das ***Fenster*** und atmet einmal tief durch.
NHH5 Für das Abendessen steht Gerda bereits seit 2 Stunden in der ***Küche*** und schält Gemüse.

NHH6 Das Schuljahr ist fast zu Ende und alle freuen sich schon auf die ***Ferien*** und ein bisschen Erholung.
NHH7 Der Schauspieler inszeniert das Stück auf der Bühne des Hamburger ***Theaters*** mit großer Bravour.
NHH8 Das Blöde ist, dass nach einem Sonntag immer ein ***Montag*** kommt und die Arbeit wieder beginnt.
NHH9 Markus hat sich große Mühe mit der Matheaufgabe gegeben, nur leider stimmt das ***Ergebnis*** nicht.
NHH10 Beim Umzug in die neue ***Wohnung*** kam es zu mehreren Zwischenfällen.
NHH11 Der Postbote wirft den ***Brief*** in den dafür vorgesehenen Briefkasten.
NHH12 Viele Geschäfte klagen, dass die Kunden nicht mehr im Laden, sondern nur noch im ***Internet*** bestellen.
NHH13 Das Brautpaar fiebert der ***Hochzeit*** schon seit Monaten sehnsüchtig entgegen.
NHH14 Von den Jahreszeiten ist mir mit seinen bunten Blättern der ***Herbst*** am liebsten.
NHH15 Das Klima im Landesinneren unterscheidet sich von jenem an der ***Küste***, welches spürbar rauer ist.
NHH16 Es fing an zu schneien und weil er fror, drehte Heinz die ***Heizung*** noch etwas weiter auf.
NHH17 Direkt nach dem Notruf rückt die Feuerwehr zu ihrem ***Einsatz*** auf dem Bahnhofsgelände aus.
NHH18 Der Kapitän ist froh, als er mit seinem Schiff endlich den sicheren ***Hafen*** erreicht.
NHH19 Das Empire State Building war eine zeitlang das höchste ***Gebäude*** New Yorks.
NHH20 Die Informationen sind zwar geheim, aber natürlich steht alles am nächsten Tag in der ***Zeitung*** und alle lesen es.

Non-anomalous sentences with low predictable target word of high frequency

NLH1 Auf der alten Kommode im Esszimmer hat seit Jahren das grüne ***Telefon*** seinen festen Platz.
NLH2 Vor vielen Jahren brachte ein Großonkel das alte ***Klavier***, aber niemand kann es spielen.
NLH3 Trotz des schlechten Wetters wurde das berühmte ***Konzert*** der Beatles in London zu einem vollen Erfolg.
NLH4 Da es morgen regnen soll, entschließt sich Lena, das ***Fenster*** heute noch nicht zu putzen.
NLH5 Nach der wilden Partynacht ist Lukas so müde, dass er in der ***Küche*** einschläft.
NLH6 Die Experten erwarten auch in diesem Sommer für die ***Ferien*** viele Staus auf den Urlaubsstrecken.
NLH7 Dem jährlichen Ausflug in das Hamburger Theater fiebert die ganze ***Klasse*** entgegen.
NLH8 Da Marie ohnehin gerne morgens arbeitet, stört es sie auch nicht, am ***Montag*** früh aufzustehen.
NLH9 Die Kundin ist überhaupt nicht zufrieden mit dem ***Ergebnis*** der wochenlangen Analysen.
NLH10 Für Geburtstagspartys und viele Gäste ist die neue ***Wohnung*** von Markus perfekt geeignet.
NLH11 Nach zwei Wochen vergeblicher Suche findet Philipp den ***Brief*** seiner Oma doch noch im Zeitungskorb.
NLH12 Viele Menschen stehen ihm zwar kritisch gegenüber, aber das ***Internet*** ist zum wichtigsten Medium geworden.
NLH13 Der emotionale Höhepunkt der ***Hochzeit*** war der Moment, in dem sich das Brautpaar das Eheversprechen gab.
NLH14 Es war ausgerechnet am Hochzeitstag, als der ***Herbst*** sich von seiner unfreundlichsten Seite zeigte.
NLH15 Zu ihrem 50. Geburtstag wünscht sich Gerda, mit der ganzen Familie an der ***Küste*** Urlaub zu machen.
NLH16 Es passiert eher selten, dass gleich in mehreren Wohnungen die ***Heizung*** ausfällt.
NLH17 Alle sind sich einig, dass die Abwehr bei ihrem ***Einsatz*** im Länderspiel völlig versagt hat.
NLH18 Schließlich entscheidet sich Karl doch noch, den sicheren ***Hafen*** der Ehe anzusteuern.
NLH19 Es gibt keine Grünfläche und keinen See in unmittelbarer Nähe des ***Gebäudes*** und der Kantine.
NLH20 Der Verteidiger sieht keine Chance, gegen die Verleumdung in der ***Zeitung*** vorzugehen.

Non-anomalous sentences with inanimate first argument

NIA1 Der Brunnen fasziniert den Beobachter wegen seiner kunstvollen Wasserspiele.
Fasziniert der Brunnen den Beobachter?
NIA2 Der Sockel gefällt dem Maler auch in der neuen Farbe noch nicht besonders.
Gefällt dem Maler der Sockel?
NIA3 Der Artikel beeindruckt den Chefredakteur dennoch sehr, trotz der vorherigen Zweifel.
Beeindruckt der Artikel den Monteur?
NIA4 Der Konsum vermittelt das Gefühl von enormer Leistungsfähigkeit und macht deshalb schnell süchtig.
Vermittelt die Leistungsfähigkeit ein Gefühl?
NIA5 Der Fluglärm ärgert den Anwohner mittlerweile kaum noch, da er sich daran gewöhnt hat.
Hat man sich an etwas gewöhnt?
NIA6 Der Bankauszug bestätigt den Verdacht auf Untreue von mehreren tausend Euro Steuergeldern.
NIA7 Der Haushaltsplan erweckt den Eindruck, dass man nur Steuern sparen wollte.
Sollte Personal eingespart werden?
NIA8 Der Generator erzeugt den Strom aus Biodiesel und speichert ihn in Akkus.
NIA9 Der Beruf verändert die Einstellung zu altbewährten Handlungsmustern und Maximen.
NIA10 Der Neubau dient dem Magistrat nicht nur zur Repräsentation der eigenen Machtansprüche.
NIA11 Der Mord verfolgt den Zeugen auch in seinen Träumen, da er so unvorstellbar grausam war.
NIA12 Der Planet durchzieht das Sonnensystem immer auf der gleichen Umlaufbahn.
NIA13 Der Flammkuchen schmeckt dem Gourmet heute gar nicht, weil er mit zu viel Petersilie garniert ist.

NIA14 Der Kampf erschöpft den Boxer zusehends, denn er dauert bereits 6 Stunden.
NIA15 Der Klimawandel beschäftigt das Parlament bereits seit Jahren, ohne dass bislang eine geeignete Lösung gefunden werden konnte.
NIA16 Der Supermarkt liefert die Lebensmittel direkt bis an die Haustüre, und das ohne Aufpreis.
NIA17 Der Pegelstand beunruhigt die Feuerwehr, da er wegen der Regenfälle immer weiter steigt.
NIA18 Der Ofen erhitzt das Wohnzimmer so stark, dass man auch bei Minustemperaturen keinen Pullover benötigt.

Anomalous sentences with animate first argument

AAA1 Der Schaffner beißt den Stein mit voller Absicht einmal fest in die Schulter.
AAA2 Der Komiker gebiert den Hering nach Ablauf von 9 Monaten Schwangerschaftszeit.
AAA3 Der Jüngling lackiert den Eintrag und verbringt anschließend den Nachmittag in der Badewanne.
AAA4 Der Komponist schnupft den Luftballon vor jedem seiner Konzerte in der Berliner Philharmonie.
AAA5 Der Fußgänger vergrault den Schrank mit seinen zahlreichen Eskapaden bei Partynächten.
AAA6 Der Drogenhändler verrät den Bildschirm an die zuständige Stabsstelle bei der Ausländerbehörde.
AAA7 Der Winzer keltert den Knoten heutzutage mit einer Prise frischer italienischer Kräuter.
AAA8 Der Patient operiert den Stuhl mit einem neuen Skalpell aus Elfenbeinholz
AAA9 Der Imker zimmert den Pfirsich aus den Resten des alten Holzschuppens zusammen.
AAA10 Der Metzger schlachtet den Baumstamm an jedem Morgen, damit das Fleisch frisch ist.
Schlachtet der Metzger den Baumstamm?
AAA11 Der Busfahrer füttert den Karton mit den Küchenabfällen vom Abendessen und schließt den Bus anschließend ab.
AAA12 Der Roboter zwingt den Gedanken dazu, auf der hohen Stadtmauer zu balancieren.
Zwingt der Roboter den Teddybären?
AAA13 Der Astronaut schmeckt das Gebäude in der Suppe, obwohl alles vorher gut erhitzt wurde.
AAA14 Der Pilot liest den Geschmack in der Freitagsbeilage der Frankfurter Allgemeinen Zeitung.
Wird etwas in der Freitagsbeilage gelesen?
AAA15 Der Urlauber überlebt den Schuh bei den Renovierungsmaßnahmen des Hauptstadtflughafens.
Wird der Hauptstadtflughafen renoviert?
AAA16 Der Klempner schneidet den Anbau des Wohnzimmers in mundgerechte Happen.
Handelt es sich um den Anbau des Schuppens?
AAA17 Das Baby vermietet das Feriendomizil in der Toskana auf der eigenen Internetseite.
Befindet sich das Domizil in Alaska?
AAA18 Die Psychologin verzehrt den Verbrennungsmotor zum Frühstück, aber abends isst sie lieber ein Wurstbrot.

Non-anomalous, high associated sentences followed by comprehension questions

NH1C Der Glückspilz gewinnt den Trostpreis bei der Tombola des Fördervereins für Suchtkranke.
Gewinnt der Glückspilz den Trostpreis?
NH2C Der Verleger liest den Roman bereits vor der öffentlichen Vorstellung auf der Buchmesse.
Liest der Verleger das Handbuch?
NH3C Der Kellner serviert den Teller bei der Gala der Behringstiftung unhöflicherweise von links.
Wird von links serviert?

Non-anomalous, low associated sentences followed by comprehension questions

NL1C Der Administrator verteilt den Link über den Gruppenverteiler an alle Mitglieder.
Wird der Link nur an ausgewählte Mitglieder verteilt?
NL2C Der Bundestag vertagt den Beschluss nach einer kurzen, aber intensiv geführten Debatte.
Vertagt der Bundestag den Beschluss?
NL3C Der Journalist entdeckt den Artikel in einer sehr angesehenen Zeitung.
Entdeckt der Buchautor den Artikel?

Anomalous, high associated sentences followed by comprehension questions

AH1C Der Schlauch flickt den Radfahrer nach der Reifenpanne mit einem Kaugummi.
Wird etwas mit einem Kaugummi geflickt?
AH2C Der Anker lichtet den Kapitän angesichts des herannahenden Gewitters hastiger als gewöhnlich.
Entfernt sich das Gewitter?
AH3C Der Honig erntet den Imker jeden Morgen im Bienenstock am Waldrand.
Erntet der Honig den Imker?

Anomalous, low associated sentences followed by comprehension questions

AL1C Der Dachfirst streicht den Maler waghalsig ohne Gerüst und nur auf einer Leiter.
Streicht der Dachfirst den Müllmann?
AL2C Der Kessel schweißt den Schlosser und feilt die Nähte anschließend glatt.
Wird etwas gefeilt?
AL3C Der Saal schmückt den Schwiegervater mit zahlreichen Girlanden und Kerzen.
Wird etwas mit Gardinen geschmückt?

Non-anomalous sentences of varying structures, followed by comprehension questions

NX1C Bei „U“ formt der Chorleiter den Mund zu einer Schnute und singt den Ton.
Macht der Chorleiter etwas?
NX2C In der katholischen Kirche ist die Verehrung von Gebeinen Heiliger und anderen Reliquien fest verankert.
Werden Gebeine von Heiligen verehrt?
NX3C Wegen der starken Hitze stinken Marias Füße in den neuen Sandalen und kleben am Fußbett.
Ist es sehr heiß?
NX4C Zum Entsetzen seiner Oma, trägt Dirk neben dem Tattoo jetzt auch ein Piercing in der Unterlippe.
Ist die Oma entsetzt?
NX5C Die Provence ist mit ihrem lila blühenden und angenehm duftenden Lavendel ein Touristenmagnet.
Blüht der Lavendel orange?
NX6C Es erfordert noch viel experimentelle Forschung, bis wir verstehen, wie die Neuronen im Gehirn unktionieren.
Ist keine weitere Forschung erforderlich?
NX7C In Museen kann man heute noch die selbstgemachten Sandalen der römischen Legionäre bewundern.
Kann man Schwerter bewundern?
NX8C Die einzelnen Tagungsteilnehmer gelten jeweils als Koryphäe auf ihrem Forschungsgebiet.
Sind die Tagungsteilnehmer Laien?
NX9C Selbst auf dem Flugplatz kann man bei guter Sicht die Insel vor der Ostküste erkennen.
Kann man etwas vor der Ostküste erkennen?
NX10C Nach den Protesten kündigt das Management der Klinik einschneidende Veränderungen an.
Sind die angekündigten Veränderungen einschneidend?
NX11C In der Küche an der Wand hängt der Kalender mit allen wichtigen Terminen.
Hängt der Kalender in der Küche?
NX12C Mit hohen Windgeschwindigkeiten fegt der Sturm über die küstennahen Gebiete.
Fegt der Sturm über die Küste?
NX13C Nach dem Abitur verlassen viele junge Leute für das Studium ihr Elternhaus.
Verlassen alle jungen Leute das Elternhaus?
NX14C Die unterschiedlichen Ansichten entfachen die Debatte zwischen den Parteien von Neuem.
Wird die Debatte nur einmal geführt?
NX15C Gerade in der Pubertät ignorieren Kinder oft den Rat ihrer Eltern und machen, was sie wollen.
Wird der Rat der Großeltern ignoriert?
NX16C Zum 25-jährigem Bestehen feiert die Firma das große Jubiläum mit einem Festakt.
Feiert der Verein ein Jubiläum?