Linguistic rhythm guides parsing decisions in written sentence comprehension

Gerrit Kentner

Institut für Linguistik, Johann Wolfgang Goethe-Universität, Grüneburgplatz 1, 60629 Frankfurt, Germany

Abstract

Various recent studies attest that reading involves creating an implicit prosodic representation of the written text which may systematically affect the resolution of syntactic ambiguities in sentence comprehension. Research up to now suggests that implicit prosody itself depends on a partial syntactic analysis of the text, raising the question of whether implicit prosody contributes to the parsing process, or whether it merely interprets the syntactic analysis.

The present reading experiments examine the influence of stress-based linguistic rhythm on the resolution of local lexical-syntactic ambiguities in German. Both speech production data from unprepared oral reading and eye-tracking results from silent reading demonstrate that readers favor syntactic analyses that allow for a prosodic representation in which stressed and unstressed syllables alternate rhythmically. The findings contribute evidence confirming immediate and guiding effects of linguistic rhythm on the earliest stages of syntactic parsing in reading.

Keywords: reading; implicit prosody; syntactic parsing; linguistic rhythm

1. Introduction

When reading silently, many readers experience an 'inner voice' that conveys from the graphemic string an intrinsic auditory version of the text. This mental representation has been described as being essentially speech-like, not only entailing segmental phonological information, but also prosody and even paralinguistic characteristics such as voice quality and speech tempo (Chafe,

Email address: gerrit@lingua.uni-frankfurt.de (Gerrit Kentner)

1988). While there is little disagreement about the existence of the 'inner voice' phenomenon, it is debated whether and how the prosodic characteristics of the implicit phonological representation affect sentence comprehension.

In the present study, we will focus on one aspect of this mental representation, namely on the linguistic rhythm that emerges from the implicit stress patterns of the word sequence. The findings of two reading experiments presented here confirm that readers mentally construct patterns of implicit lexical prominences, which evolve from the concatenation of individual words. In the face of a temporal syntactic ambiguity, readers preferably generate a parse that conforms to rhythmic well-formedness principles. Specifically, the findings indicate that the initial stages of the syntactic parsing process are sensitive to the local prosodic environment even in the written modality, where no explicit prosodic cues exist.[1]

In the following section, we will briefly review findings on the relevance of linguistic rhythm in auditory language processing and discuss existing research on the role of implicit prosody for written sentence comprehension. Together, these findings motivate the two reading experiments, which are designed to shed light on the interplay of linguistic rhythm and syntactic parsing in reading.

1.1. Stress and linguistic rhythm in auditory language

Stress is hierarchical in the sense that for each content word there is a single syllable that carries the main stress (Hayes, 1995). Other syllables within the word either bear secondary stress or remain unstressed. Although there is no clear phonetic correlate of stress, stressed syllables are usually lengthened and may be realized with a higher pitch compared to unstressed syllables (Beckman and Pierrehumbert, 1986; Hayes, 1995). The hierarchical nature of stress also implies that different levels of stress have to be distinguished for different prosodic domains. Lexical stress determines the prominent syllable within a word. Beyond the lexical level, the prosodic phrase carries prominence on one of its lexical constituents, which may be realized as a pitch accent on the stressed syllable. Likewise, among phrases within a sentence, one is assigned the nuclear accent, realized on the stressed syllable of the most prominent word within that phrase. The assignment of

[1]Under certain circumstances, commas might serve as cues to prosodic phrasing (Steinhauer and Friederici, 2001; Steinhauer, 2003). The reliability of the comma-prosody correlation, however, crucially depends on the context (Chafe, 1988).

phrasal and sentence stress is mainly determined by the syntactic structure and the discourse context (Gussenhoven, 1983; Selkirk, 1995; Truckenbrodt, 2006). Word stress in German is lexically specified since it is not completely predictable from the segmental and syllabic structure of the word (Wiese, 2000).

Languages like German and English exhibit a general preference for an alternation of strong (i.e. stressed) and weak syllables, which manifests itself particularly in the avoidance of stress clashes, i.e. the avoidance of two adjacent syllables carrying main word stress. Stress clash avoidance has been demonstrated to affect language production in various ways: speakers might deviate from the citation form stress pattern to resolve a potential stress clash (Hayes, 1995; Selkirk, 1984); when faced with visually presented pseudo-words in sentential context, speakers have been shown to favor a stress pattern that maximizes rhythmic alternation (Kelly and Bock, 1988). The preference for rhythmic alternation may also have syntactic consequences for language production: given the choice, speakers preferably use syntactic constructions that prevent a stress clash (Anttila et al., 2010; Schlüter, 2005; Speyer, 2010).

As for auditory language comprehension, listeners were shown to be sensitive to rhythmic regularity in speech. Dilley and McAuley (2008) and Niebuhr (2009) report that listeners analyze the same lexically ambiguous syllable sequence differently depending on the linguistic rhythm (trochaic or iambic) established by the preceding context. Using event related potentials (ERP), Schmidt-Kassow and Kotz (2009) showed that listeners are sensitive to deviations from trochaic speech patterns when explicitly asked to judge the rhythmicity of the stimulus sentences. Niebuhr (2009) proposes that the phonetic rhythm has a guiding function in speech perception, in that it makes upcoming material predictable.

Warren et al. (1995) show that stress patterns on critical words have the potential to impinge on the syntactic analysis of temporarily ambiguous sentences. Specifically, their findings suggest that the perception of stress shift on critical words augments the cues to upcoming phrase boundaries even before such a boundary is encountered.

1.2. The generation of prosody in reading

Skilled readers produce prosody in accordance with the syntactic structure (Koriat et al., 2002) and also with the information structural analysis of the text. These factors especially influence accentuation and prosodic phrasing, implying that reading aloud simultaneously involves syntactic parsing, the interpretation of context, and the production of accordant prosody. The involvement of prosody in silent reading is less obvious, especially given the lack of a clear correlate of prosody in written text.

Recent research by Ashby and colleagues (Ashby and Clifton, 2005; Ashby and Martin, 2008; Ashby and Rayner, 2004) verifies the involvement of prosodic processing in silent reading on the lexical level. Ashby and Clifton (2005) demonstrate that readers fixate words with two stressed syllables (*situation*) longer than words with only one stressed syllable (*authority*), irrespective of the word length and frequency. Employing eye-tracking and ERP, Ashby and Martin (2008) find that readers routinely activate a prosodic phonological representation of the lexical items within the first 100 ms upon visual encounter. Ashby and Martin (2008) take this as evidence for an early speech-like phonological representation of the text being read.

The notion of speech-likeness suggests that the implicit prosody generated in the reading process is not to be understood as a simple concatenation of lexical prosodic structures. Instead, speech prosody is supralexical in nature, a condition that is evidenced, for example, by the stress shift phenomenon. To put it differently, if implicit prosody were speech-like, it should be subject to conditions of linguistic rhythm and the preference for an alternation of strong and weak syllables. Direct evidence for effects of linguistic rhythm in silent reading, however, is currently missing.

1.3. The role of implicit prosody in written sentence comprehension

Since prosody is not explicitly encoded in the graphemic string, its role in written sentence comprehension has been controversial: it is unclear whether the prosodic representation only reproduces the syntactic analysis by the reader (Kondo and Mazuka, 1996; Koriat et al., 2002) or whether implicit prosody itself contributes to the syntactic analysis during written sentence comprehension (Bader, 1998; Fodor, 1998, 2002).

A number of studies indicate that the silent prosody readers impose on the written text does affect the syntactic analysis. Bader (1998) finds that syntactically ambiguous sentences induce stronger processing difficulties in reading when the competing syntactic structures differ with respect to their prosodic features. He proposes the *Prosodic Constraint on Reanalysis* stating that revising a syntactic structure is particularly difficult if it necessitates a concomitant reanalysis of prosodic structure. Bader (1998) substantiates this proposal with reading data on temporarily ambiguous structures, the readings of which differ with respect to accent placement. Breen and Clifton (2011) examine the processing of lexical stress on noun-verb homographs (<u>pre</u>sent – pre<u>sent</u>) in syntactically ambiguous structures. Their results suggest that the reanalysis of lexical stress aggravates the resolution of syntactic ambiguities in silent reading.

Other studies focus on the effect of phrase length in relation to syntactic attachment preferences. Hirose (2003), Hwang and Schafer (2009) and Hwang and Steinhauer (2011) found that readers posit syntactic clause boundaries in temporarily ambiguous sentences based on the length of the preceding constituent. This leads to reading difficulties if the boundary turns out to be incompatible with the upcoming material. Several studies underpin the implicit prosodic effect in silent reading with consistent data obtained from oral reading experiments (Augurzky, 2006; Hirose, 2003; Hwang and Schafer, 2009; Jun, 2003). Others, however, fail to find the predicted correlation of attachment preference and overt prosodic pattern (Bergmann et al., 2008; Jun, 2010).

In summary, the research reviewed here clearly favors an account which grants implicit prosody a functional role in written sentence comprehension. As to the question of when and how exactly prosodic processes constrain the syntactic analysis in reading, the research on implicit prosody so far suggests that at least a partial syntactic analysis of the critical words and phrases is required in order for implicit prosody to show its effects on written sentence comprehension. Augurzky (2006) concludes from a thorough review and her own ERP data that "the parser initially leaves the prosodic analysis underspecified" (p. 206). Accordingly, prosodic effects on interpretation in silent reading would only occur in a very late processing stage.

Other studies emphasize the immediate nature of the prosodic effect. Recently, Hwang and Steinhauer (2011) presented ERP evidence suggesting that relatively long phrases trigger the insertion of implicit prosodic boundaries; apparently, the processing mechanism immediately interprets the implicit prosodic break to signal a syntactic phrase boundary. Note, however, that, in order to evaluate the length of phrases, the processing mechanism has to merge several words to form such phrases in the first place.

Correspondingly, existing research on implicit prosody is consistent with the idea that reading prosody depends on a partial syntactic analysis of the text. This generalization, however, might be due to the fact that the experiments mostly scrutinize effects of larger prosodic domains (prosodic phrasing, phrasal accentuation) on syntactic parsing. More local prosodic features like lexical stress and linguistic rhythm have, as yet, been largely disregarded. These factors may, however, more directly affect the assignment of syntactic structure. Clearly, different prosodic features might have different effects on the process of reading comprehension.

The work by Ashby and colleagues (Ashby and Clifton, 2005; Ashby and Martin, 2008) suggests that (lexical-)prosodic information such as syllable structure and stress pattern of the words is available to the processing mechanism from very early on in reading. It would be astonishing if it were not used immediately, especially since such information may be meaningful for the comprehension process (cf. Dilley and McAuley, 2008; Niebuhr, 2009; Warren et al., 1995). Under the assumption of a speech-like prosodic-phonological representation in reading, and given their immediate availability, stress and linguistic rhythm should exert their influence from the very beginning of the parsing process.

The following experiments are designed to put this hypothesis to a test and to show that even the earliest steps of syntactic parsing (i.e. the determination of the syntactic category of an ambiguous lexical item) may be guided by the implicit rhythm that emerges from the stress patterns readers impose on the written words.

2. Experiments

Given the general preference for the alternation of strong and weak syllables in German, it is predicted that a stress clash is avoided wherever more rhythmic alternatives are available. Despite the lack of explicit encoding of stress in written text, this should be true for reading aloud as well as for silent reading if readers indeed generate a speech-like phonological representation as proposed by Ashby and Martin (2008). This has consequences for the syntactic processing of the sentence: in the face of an ambiguous structure that involves a stress clash in one reading but not in the other, there should be a preference for the version without stress clash. This hypothesis will be tested in two reading experiments.

The object of investigation is syntactically ambiguous structures like (1), the two readings of which are differentiated prosodically by accentuation (stressed syllables underlined, accented syllables in capital letters).

(1) Der Polizist sagte, dass man...
The policeman said that one ...

a. ... nicht mehr NACHweisen kann, wer der Täter war. TEMP-INI
... couldn't prove anymore who the culprit was.

b. ... nicht mehr erMITteln kann, wer der Täter war. TEMP-MED
... couldn't determine anymore who the culprit was.

c. ... nicht MEHR nachweisen kann, als die Tatzeit. COMP-INI
... couldn't prove more than the date of the crime.

d. ... nicht MEHR ermitteln kann, als die Tatzeit. COMP-MED
... couldn't determine more than the date of the crime.

In (1), two different syntactic analyses of *mehr* are reflected in different prosodic renderings. In (1-a) and (1-b), *mehr* is part of the temporal adverbial *nicht mehr* (TEMP) and remains unaccented.[2] In this case, the following verb receives the main phrase accent. In (1-c) and (1-d), *mehr* is a comparative quantifier (COMP) that serves as a complement to the verb. In its function as complement to the verb, *mehr* receives main phrase accent, i.e. it is marked by a rising pitch accent. When preceded by an accented complement as in (1-c) and (1-d), the verb typically need not bear an accent (Truckenbrodt, 2006).

Since accent information is not encoded orthographically, the sentences are disambiguated in written text only after the verb complex, i.e. in the phrase that closes the sentence. In the temporal reading, the disambiguating phrase is a sentential argument of the verb that follows the ambiguous *mehr*. In the comparative reading, *mehr* itself is the complement of the verb and the disambiguating phrase is the extraposed comparative complement of *mehr* introduced by the standard marker *als* (engl. *than*).[3]

[2]The semantics of the lexical unit *nicht mehr* in the temporal adverbial sense cannot be analyzed compositionally. It is therefore questionable whether the graphemic word *mehr* has an independent lexical status in this context.

[3]According to German comma rules, the sentential complement in (1-a) and (1-b) (temporal reading) is separated by a comma. As for the comparative reading, a comma is

A similar syntactic ambiguity involving *mehr* was studied by Bader (1996) in a self-paced reading experiment. His results suggest that, generally, the temporal, unaccented reading of *mehr* is preferred over the comparative, accented reading, which Bader attributes to a general avoidance of (implicitly) accenting function words. Another possible source of this preference might be a higher frequency of the temporal *mehr*-construction.

For the purpose of this experiment, the rhythmic environment is systematically varied at the verb following *mehr*.[4] The verb has either initial stress (INI) as in (1-a) (TEMP-INI) and (1-c) (COMP-INI) or medial stress (MED) as in (1-b) (TEMP-MED) and (1-d) (COMP-MED). Condition COMP-INI, as opposed to all other conditions, involves a stress clash that is brought about by the adjacency of accented *mehr* and a verb with initial stress. The effect of this rhythmic imperfection on syntactic parsing will be tested in oral reading (experiment I) and in silent reading (experiment II).

3. Experiment I

A speech production experiment was set up to test the influence of the rhythmic environment on the resolution of the local syntactic ambiguity concerning the word *mehr* in sentences like (1) in oral reading. The experiment consisted of two sessions in direct succession. In the first session, the 'unprepared session', participants read the stimuli out loud without advance preparation, i.e. without having knowledge of the disambiguation prior to executing the task. This way, the realization of accent on the critical word *mehr* should reflect the initial analysis unaffected by the disambiguating context. In the second session, the 'prepared session', participants were asked to familiarize themselves with the complete sentences before reading them out aloud.

3.1. Materials

24 sets of sentences like (1) were devised that contain a local syntactic ambiguity in writing but are unambiguous when spoken because of relevant prosodic cues. The actual sentences used in the experiments are listed in

---

required only if the *als*-phrase is a clause, i.e. if it features an overt main verb. Although two-thirds of the comparative items in this experiment do not occur with a sentential disambiguating phrase but with an NP, the comma is set throughout to ensure comparability across conditions.

[4]Bader (1996) did not investigate rhythmic effects.

the appendix. All critical verbs following *mehr* are obligatorily transitive verbs that can take an NP or a sentential object to satisfy their argument structure requirements. The critical verbs are all trisyllabic, prefixed verbs that appear in their infinitival form and precede an inflected modal verb.

3.1.1. Validation of materials

The 12 verbs with initial stress and the 12 verbs with medial stress were matched with respect to word-form frequency and length. Word form frequencies were obtained from the Leipzig Wortschatz corpus (http://wortschatz.uni-leipzig.de/), which consists of approximately 50 million sentences of German newspaper text collected between 1994 and 2008. The mean logarithmized frequency is 7.14 (1.075 standard deviation) for verbs with initial stress and 6.93 (1.3) for verbs with medial stress. A linear model that evaluates the word frequency against the verb type does not suggest any significant difference between the two types ($F=0.186$, $df=22$, $p=0.67$).

Since the sentences with initial versus medial verb stress (examples in (1)) differ not only with respect to the verbal stress pattern but – necessarily – also with respect to the semantics of the verb (despite some effort to choose semantically similar verbs), a validation of the comparability of the conditions is required. To this end, all items were subjected to a sentence rating study.

The experimental sentences were distributed over four lists using a Latin square design with conditions counterbalanced across lists. In this way, each list presented 24 experimental sentences, six from each condition. In each list, the experimental sentences were interspersed with 76 filler sentences from four unrelated experiments. The order of the items was pseudo-randomized using the Mix randomization tool by van Casteren and Davis (2006) such that items from the same experiment had a minimum distance of three and items from the same experimental condition had a minimum distance of six. Each list was printed on A4 paper in landscape layout with each sentence presented on a single line.

Forty-six first-year undergraduate students from the University of Potsdam, all naïve to the purpose of the experiment, took part in the rating study for course credit or payment. They were each given one of the four lists. The subject's task was to rate every sentence on a seven point Likert scale (1 - easy and perfectly acceptable sentence – 7 - incomprehensible, unacceptable sentence) and note the respective number next to each sentence on the sheet. No time constraints were given. All participants completed the rating task within 40 minutes.

Of the total 1104 sentences, 48 (4%) had missing or unidentifiable ratings. The 1056 obtained ratings were treated as numerical values. The boxplot in Figure 1 depicts the median and the distribution of the ratings by condition.

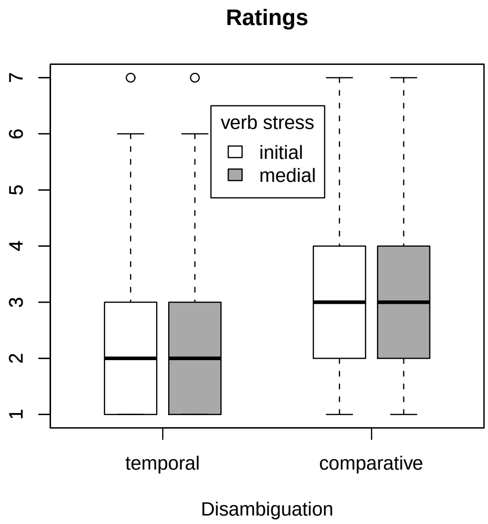

Figure 1: Boxplots representing the distribution of ratings broken down by 'disambiguation' and 'verb stress'.

The ratings were evaluated against the crossed fixed factors 'disambiguation' (with the two levels TEMP and COMP) and 'verb stress' (with the two levels INI and MED) using a linear mixed model (Bates and Sarkar, 2007; Gelman and Hill, 2007). Participant and item were treated as random variables. Table 1 summarizes the results of the model.[5] The model reveals a significant main effect for 'disambiguation'. The effect for 'verb stress' and the interaction are non-significant. The rating results thus do not indicate any difference in terms of acceptability of the sentences that is systematically attributable to the implicit rhythmic environment brought about by the stress pattern of the verb. However, the significant main effect of 'disambiguation' shows that the temporal reading of *mehr* (TEMP) is on the whole more acceptable than the comparative versions (COMP).

[5]Since, in linear mixed models, determining the precise degrees of freedom is non-trivial, the t-values are approximations. An absolute t-value of 2 or greater indicates statistical significance at $\alpha = 0.05$.

Coefficient	Estimate	Std. Error	t-value
disambiguation	0.2696	0.0374	7.207
verb stress	0.0077	0.0375	0.206
disamb × v-stress	0.0392	0.0374	1.048

Table 1: Results of linear mixed model evaluating the ratings against the crossed fixed factors ‘disambiguation’ and ‘verb stress’.

3.2. Experimental procedure

For the oral reading experiment, the experimental sentences were again distributed over four lists with conditions counterbalanced across lists. In each list, the 24 experimental sentences were embedded in 69 filler sentences from four unrelated experiments. The total of 93 items was pseudo-randomized for each subject independently, such that sentences from the same experiment had a minimum distance of three items and sentences from the same experimental condition had a minimum distance of eight items. Participants saw the same list of items in the same order in both unprepared and prepared sessions of the experiment.

The experiment took place in an anechoic room with an AT4033a Audio-Technica studio microphone. Each participant was seated in front of a 15” computer screen with the microphone placed approximately 30 cm from the participant’s mouth. A keyboard was placed on a table in front of the subject. Recordings were made on a computer using the RecordVocal function of the DMDX (Forster and Forster, 2003) and a C-Media Wave sound card at a sampling rate of 44.1 kHz with 16 bit resolution.

Each of the two sessions was preceded by three example stimuli (not related to any of the experimental stimuli) for the participants to familiarize themselves with the task.

For the unprepared session, the DMDX presentation was programmed for each item as follows: First, only the first one or two words (the sentence initial subject noun phrase or proper name) were presented on the screen. Participants were told to familiarize themselves briefly with these words. They were instructed to then press the space bar, inducing the presentation of the entire sentence. Participants were asked to start reading the sentence out loud immediately as it appeared on the screen and to do so as fluently as possible. Pressing the space bar automatically initiated the recording. After a fixed recording time of five seconds, the procedure was repeated for the next item. For each sentence, only one realization per subject was recorded.

No corrections were recorded in the case of hesitations or slips of the tongue.

After completion of the unprepared reading session, participants were encouraged to take a short break of approximately five minutes, which was followed by the prepared session. The item presentation differed from the unprepared session in that readers were presented with the whole sentence from the start and were told to familiarize themselves with the sentence before reading it out loud. Again, when ready to read out loud, readers were asked to press the space bar to initiate the recording. This time, pressing the space bar did not change the visual presentation. For each item, the recording time was set to five seconds, after which the next item appeared on screen. For each item, again only one realization per subject was recorded.

3.2.1. Participants

Twenty-four female first-year undergraduate students from the University of Potsdam took part in the experiment. All were native speakers of German and naïve as to the purpose of the experiment. They either received course credit or were paid 5 Euros for their participation.

3.3. Predictions

In unprepared reading, i.e. if readers are unaware of the disambiguation, accentuation of *mehr* should be avoided given that the unaccented, temporal reading is processed more easily and is generally preferred (c.f. Bader (1996) and section 3.1.1.). Moreover, the predicted avoidance of stress clash is hypothesized to lead to a higher number of unaccented realizations of *mehr* in the case of a following verb with initial stress.

On the assumption that successful reading necessitates conformity of prosodic and syntactic structure, realizations of *mehr* that are prosodically incompatible with the disambiguating region should lead to reading difficulties – that is, readers might be led down the garden path if their prosodic realization of *mehr* turns out to be infelicitous. Since reading comprehension is strongly correlated with reading fluency, the difficulties should manifest themselves in hesitations or a slowdown in speech (e.g. Fuchs et al., 2001) once the reader reaches the disambiguating region.

As for prepared reading, the disambiguation is known to the reader before oral realization. The disambiguation reveals the lexical-syntactic status and with it the appropriate accent for *mehr* – it should thus be the decisive factor for the accentuation of *mehr*. The immediate rhythmic environment does not alter the grammatical requirement of accentuation on *mehr* and therefore should not have a systematic effect.

3.4. Data analysis

All in all, 1152 experimental sentences were recorded, 576 in the unprepared session and 576 in the prepared session. The sentences from the two sessions were independently judged by two students each. The judges were not informed about the conditions and the purpose of the experiment before completion of their job. Their task was i) to note slips of the tongue and disfluencies in the part of the sentence up to but excluding the disambiguating phrase, and ii) to determine for each sentence if the word *mehr* was accented or not, i.e. if it was to be understood as a comparative complement or as a temporal adverbial. In order to avoid an influence of the disambiguating region on the judgments, all sound files were cut after the verb complex prior to the judgment process. The sentences were presented to the judges in randomized order. The judges were paid for their work.

For ease and clarity of exposition, the results of the prepared reading task will be reported before the results of the unprepared reading session.

3.5. Results

3.5.1. Results for prepared reading

Twenty-four (4%) of the total 576 sentences were marked by at least one of the judges as non-fluent or containing slips of the tongue in the region preceding the disambiguating phrase. A generalized linear mixed model (GLMM) with a binomial link function (Bates and Sarkar, 2007; Gelman and Hill, 2007; Quené and Van den Bergh, 2004) was fitted to check whether the distribution of flawed sentences is related to the experimental factors. The fixed factors of this model were i) 'disambiguation' (COMP vs. TEMP) and ii) 'verb stress' (INI vs. MED) with flawed versus fluent realization as the dependent variable; participant and item were included as random effects (grouping variables). Orthogonal contrast coding was applied (factor 'disambiguation': comparative=1, temporal=-1; factor 'verb stress': initial=1, medial=-1). This model does not reveal any systematic influence of the controlled factors 'disambiguation' and 'verb stress' or their interaction on the distribution of flawed sentences (all z-values $<|2|$, all p-values $>.2$). As for the 552 fluent sentences, the assessments of the two judges concerning the accentuation of *mehr* concur in 532 cases (97%). The bar plot in Figure 2 shows the percentage of accented *mehr* by condition for the consistently judged sentences. The target word was perceived as accented in the comparative readings (COMP-INI, COMP-MED) in around 90% of the cases; as for the temporal reading (TEMP-INI, TEMP-MED), *mehr* was perceived as accented in less than 10% of the cases.

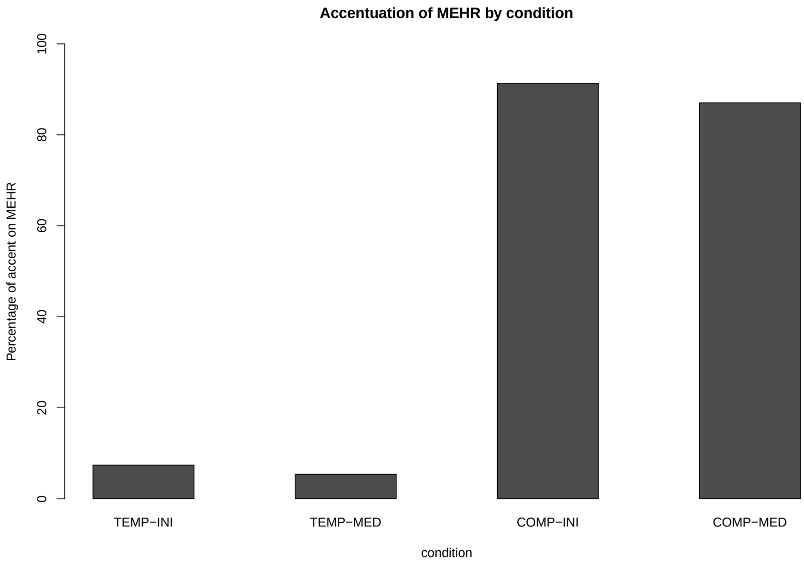

Figure 2: Accentuation of *mehr* as determined by judges broken down by condition in prepared reading experiment.

The accentuation status of *mehr* was again evaluated with a GLMM incorporating the same fixed factors and grouping variables as above. In line with the above predictions, this model confirms a single significant main effect for the fixed factor 'disambiguation'. The main effect for 'verb stress' and the interaction remain non-significant (cf. Table 2).

Coefficient	Estimate	Std. Error	z-value	p-value
disambiguation	2.96412	0.20989	14.122	<0.001
verb stress	0.26280	0.18773	1.400	0.162
disamb × v-stress	0.04809	0.18752	0.256	0.798

Table 2: Results of GLMM on accentuation of *mehr* in consistently judged sentences in prepared reading experiment.

3.5.2. Results for unprepared reading

In the unprepared session, 63 sentences (11%) were non-fluent or contained slips of the tongue in the region preceding the disambiguating phrase, as determined by at least one of the judges. As for the accentuation status of *mehr*, the judges agreed on 495 of the 513 fluent sentences (96%). The 495 consistently judged sentences were hand-annotated by a phonetically trained student who was blind to the purpose of the experiment and to the judgments of her fellow students. For each of the 495 sentences, the critical region starting with *nicht* up to the end of the verb complex was segmented into words and syllables and labeled accordingly.

Flawed sentences

The number of flawed sentences is relatively high (n=63, 11%), which can be partly explained by the task (unprepared reading) and the length of the sentences (10 words up to the disambiguating region). It was checked whether the distribution of flawed sentences is systematically related to the controlled factors of the experiment. No significant effect was found for either of the fixed factors ('disambiguation': *z=-0.501, p=0.62*; 'verb stress': *z=-0.747, p=0.46*), or the interaction (*z=1.017, p=0.31*), suggesting that the controlled variables do not systematically influence the distribution of flawed sentences.

Judgments on realizations of mehr

The bar plot in Figure 3 displays the percentages of accented *mehr* as perceived by the judges in the four conditions. In total, *mehr* was perceived as accented in about 24% of the cases.

Speakers accented *mehr* in 20% of the sentences with comparative disambiguation. Twenty-seven percent of the occurences of *mehr* were judged as accented in sentences with temporal reading. When the verb following *mehr* has medial stress, speakers accented *mehr* in 28% of the sentences, compared to 19% when the verb has initial stress.

A GLMM was fitted with perceived accentuation of *mehr* as the dependent variable. The fixed factors of this model are again i) 'disambiguation' (TEMP vs. COMP) and ii) 'verb stress' (INI vs. MED). Speakers and items served as random effects. This model (Table 3) yields a significant main effect for the stress position on the verb. The effect of the disambiguating region is significant, too. The interaction of stress position and disambiguation is not significant.

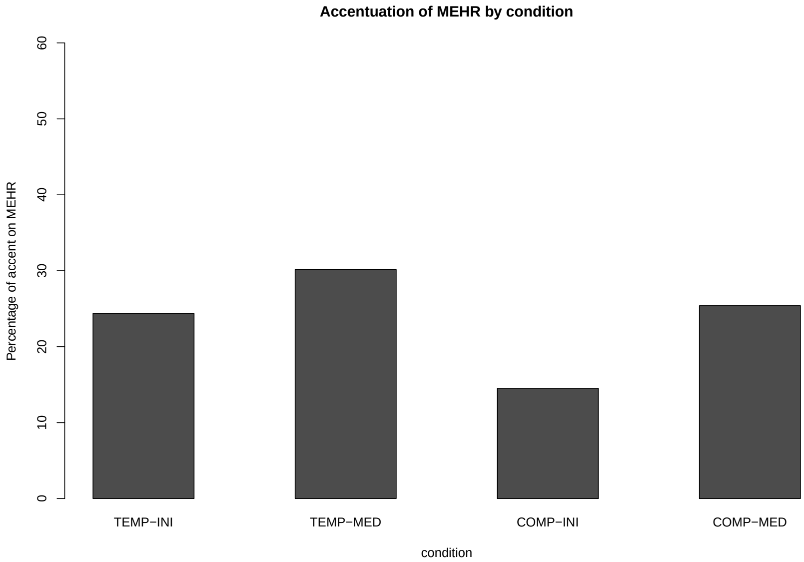

Figure 3: Accentuation of *mehr* as determined by judges broken down by condition in unprepared reading experiment.

Coefficient	Estimate	Std. Error	z-value	p-value
disambiguation	-0.2613	0.1258	-2.077	0.038
verb stress	-0.3601	0.1262	-2.853	0.004
disamb × v-stress	-0.1351	0.1254	-1.077	0.282

Table 3: Results of GLMM on perceived accentuation of *mehr* in consistently judged sentences in unprepared reading experiment.

A comparison with the prepared reading data reveals that the accentuation status of *mehr* is frequently inappropriate relative to the subsequent disambiguation. In conditions COMP-INI and COMP-MED in particular, only 20% of the trials were realized with the required accent on *mehr*. In contrast, *mehr* congruously remained unaccented in the temporal conditions TEMP-INI and TEMP-MED in 72% of the cases. Given the abovementioned main effect of verb stress on the realization of accent, the avoidance of accent on *mehr* in the comparative conditions should result in even more instances of realizations that are incompatible with the disambiguating region when the verb features initial stress. To check for this interaction, a GLMM was fit. The dependent variable this time was the appropriateness of accentuation relative to the disambiguating region. The model confirms a clear main effect of 'disambiguation' and reveals that the interaction between 'disambiguation' and 'verb stress' is significant (cf. Table 4).

Coefficient	Estimate	Std. Error	z-value	p-value
disambiguation	-1.2057	0.1090	-11.060	<0.001
verb stress	-0.1007	0.1090	-0.924	0.3557
disamb × v-stress	-0.2471	0.1090	-2.267	0.0234

Table 4: Results of GLMM evaluating the compatibility of accentuation of *mehr* relative to the disambiguating region.

Phonetic analysis of accented vs. unaccented realizations

Overall, perceived accentuations of the target word are conspicuously rarer in unprepared reading as compared to prepared reading. This is most likely due to the general preference for the unaccented, temporal reading that was attested by Bader (1996) and confirmed in the sentence rating study above.

In order to exclude misperception by the judges, their assessment was validated by means of a phonetic analysis. Also, since listeners may perceive prominence patterns on syllable sequences in context even in the absence of definite acoustic cues for such a pattern (Dilley and McAuley, 2008), a validation of their judgments is appropriate. Hence, the syllable durations and pitch contours of sentences with perceived accented and unaccented *mehr* were compared. Specifically, the region starting with *nicht* up to the modal verb preceding the disambiguating phrase was analyzed.

The upper panel of Figure 4 shows the grand average pitch contours in the critical region broken down by accentuation of *mehr* and the stress position on the following verb. The pitch contours were created by dividing each syllable in the region of interest into three equal-sized intervals and interpolating the normalized mean F0 for each of these intervals; the normalization factor used is the inverse of the maximum F0 of each sentence. The bar plots in the lower panel display the respective average syllable durations in milliseconds. Clearly, the tokens of *mehr* that were perceived as accented display longer durations compared to unaccented tokens. Moreover, there is a clear rising pitch contour on *mehr* in the accented versions (black lines), indicating the realization of a pitch accent on this word. The versions with unaccented *mehr* (grey lines) show falling pitch on the critical word and the rise appears only later on the stressed syllable of the following verb, which, in these cases, carries the phrase accent. The accentuation of *mehr* appears to already have small effects on the duration and pitch contour of the preceding *nicht* and continues to have durational effects on the realization of the following verb. Irrespective of accentuation on *mehr*, the modal verb ends on a relatively high pitch, indicating a continuation rise preceding the disambiguating phrase.

A linear mixed model with subject and item as random effects confirms a significant effect of perceived accentuation on the prosodic rendering of *mehr*. The dependent variable of this model is the pitch slope on *mehr*, i.e. the difference between the F0 values at the onset and the offset of *mehr* divided by the duration of *mehr*. The perceived accentuation serves as the fixed effect, yielding a coefficient estimate of 66.36 with a standard error of 4.88 (*t-value= 13.60*). The phonetic analyses confirm the prosodic difference between perceived accented and unaccented versions and thus validate the judgments.

Phonetic analysis of garden path effect

On the assumption that the realization of accent on *mehr* conforms to the syntactic analysis, the readers/speakers should experience comprehension difficulties when the realization of *mehr* is incompatible with the disambiguating region. In fluent oral reading, the reader's eyes are a few words ahead of the voice; hence, the slowdown in speech should already be observable at the beginning or even before the disambiguating region is spoken aloud. Correspondingly, the modal verb and the pause preceding the disambiguating phrase might be affected by the slowdown and show longer durations when the disambiguating region is inappropriate relative to the accentuation applied on *mehr*. To test for this garden path effect, the durations of both the modal verb and the pause preceding the disambiguating phrase were summed and evaluated. Specifically, the duration from the onset of the modal verb up to the onset of the disambiguating phrase was measured. By inclusion of the modal verb, effects of final lengthening due to the clause break are included in the analysis.

The boxplots in Figure 5 depict the distribution of the clause break durations (modal verb plus pause) in each condition broken down by the compatibility of the accentuation applied by the reader relative to the disambiguation. When the speakers' realization of *mehr* is inappropriate relative to the disambiguating phrase (incompatible realizations), the duration data indicates a marked slow-down compared to the appropriate realizations. A

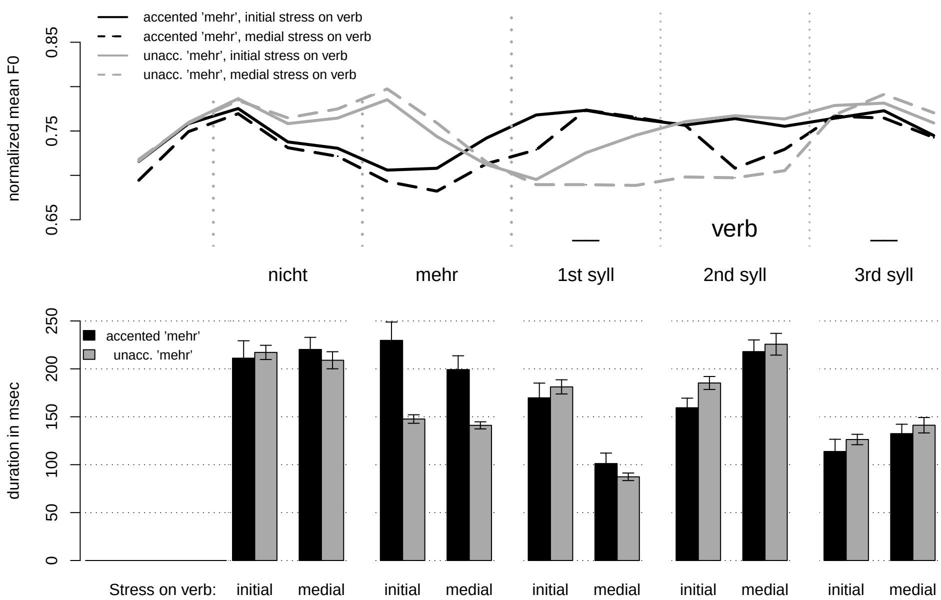

Figure 4: Grand average pitch contours (upper panel) and durations (lower panel) for each syllable in the region starting with *nicht* up to the critical verb broken down by realization of accent and verbal stress pattern.

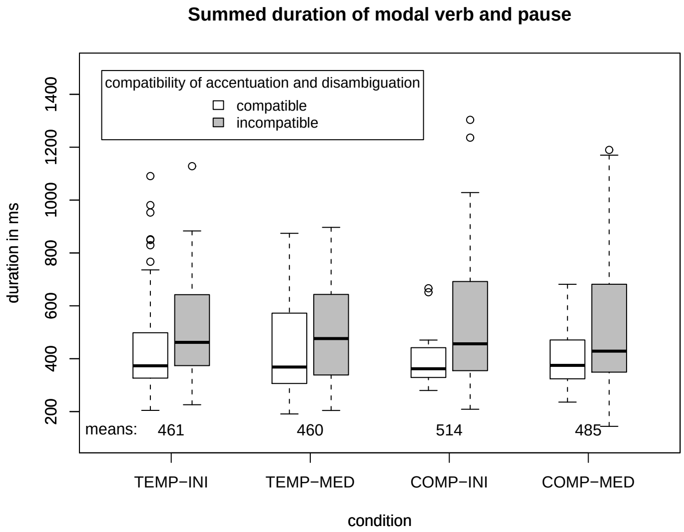

Figure 5: Duration of modal verb and following pause before the disambiguating clause in the four conditions broken down by compatibility of accentuation.

linear mixed model with the fixed factors 'compatibility of realization', 'disambiguation', and 'verb stress' and subjects and items as random effects yields a significant main effect for 'compatibility' on duration, confirming that speakers indeed slow down when the disambiguating region does not conform to the accentuation of the ambiguous *mehr*. The effect of 'disambiguation' is not significant, nor is the effect of 'verb stress' or any of the interactions. The parameters of this model (cf. Table 5) thus suggest that the compatibility of accentuation has similar effects irrespective of the presented condition.

Note, however, that some of the factors of the model are highly correlated: due to the preference for the unaccented temporal reading of *mehr*, significantly more compatible realizations were made in conditions TEMP-INI and TEMP-MED as compared to COMP-INI and COMP-MED; moreover, more compatible realizations were produced in condition COMP-MED compared to the clash condition COMP-INI. A second analysis (cf. Table 6) evaluates the duration data against the fixed factors 'disambiguation' and 'verb stress' only, thereby avoiding any correlation. This model reveals a significant main effect of 'disambiguation'.

Although the interaction does not reveal a significant effect, closer in-

Coefficient	Estimate	Std. Error	t-value
compatibility	-0.0901	0.0164	-5.48
disambiguation	-0.0094	0.0162	-0.58
verb stress	0.0078	0.0162	0.48
comp × disamb	-0.0208	0.0188	-1.11
comp × v-stress	-0.0175	0.0164	-1.07
disamb × v-stress	-0.0017	0.0164	-0.11
comp × disamb × v-stress	-0.0127	0.0165	-0.77

Table 5: Results of linear mixed model evaluating the summed duration of modal verb and pause against compatibility of accentuation, disambiguation and verb stress.

Coefficient	Estimate	Std. Error	t-value
disambiguation	0.0412	0.0137	3.02
verb stress	0.0197	0.0137	1.45
disamb × v-stress	0.0166	0.0137	1.21

Table 6: Results of linear mixed model evaluating the summed duration of modal verb and pause against the factors disambiguation and verb stress.

spection shows that the stress clash condition COMP-INI leads to significantly longer durations compared to COMP-MED (*coeff. estimate=0.0355, std.err.=0.0169, t-value=2.10*). In contrast, the durational difference between the two temporal conditions TEMP-INI and TEMP-MED does not appear to be systematic (*coeff. estimate=-0.0005, std.err.=0.0203, t-value=0.02*).

In summary, the analysis concerning the duration of the modal verb plus the pause preceding the disambiguating region presents evidence that incompatible realizations of accent on *mehr* lead to a garden path effect, indicating that speakers made a syntactic commitment when choosing the accent status of *mehr*. On average, durations of the region of interest were longer in the comparative reading (COMP-INI and COMP-MED) compared to the temporal versions (TEMP-INI and TEMP-MED), with the longest durations found in the clash condition COMP-INI. Since the increase in duration due to inappropriate accentuation appears to be similar across conditions (cf. Figure 5 and Table 5), the number of compatible vs. incompatible accentuations in the four conditions is most likely responsible for the difference in mean duration between conditions.

3.6. Discussion

The accentuation patterns of the target word in the prepared reading session conform to expectations: the ambiguous item *mehr* is accented when used as a comparative (engl. *more*) but remains unaccented in the temporal reading of *nicht mehr* (engl. *not anymore, no longer*). That is, if readers have full access to the disambiguating material before starting to read out loud, they audibly use accentuation to signal the appropriate variant of *mehr*. The immediate rhythmic environment (the verb stress manipulation) does not systematically contribute to the accentuation status of *mehr* in the prepared reading session. This also fits the expectations according to which the requirement for the accentuation of *mehr* is solely driven by its syntactic status.

As for unprepared reading, readers chose to accent the critical word *mehr* in just under 25% of the cases, indicating a preference for the unaccented, temporal version. This effect conforms to the findings by Bader (1996) and the rating study (3.1.1), which showed higher acceptability of the temporal conditions TEMP-INI and TEMP-MED as opposed to conditions COMP-INI and COMP-MED. Importantly, the judgments concerning the accentuation of *mehr* in unprepared reading reveal a significant main effect for the verbal stress pattern on the realization of *mehr*, confirming that the accentuation of the target word is systematically influenced by the immediate rhythmic environment: as hypothesized, speakers avoid accenting *mehr* when this would induce a stress clash configuration with the following verb. As predicted, this rhythm-induced avoidance of accent leads to a significantly higher number of inappropriate realizations in the context of the comparative disambiguating region.

Unexpectedly, accent on *mehr* was realized significantly more often in temporal versions (TEMP-INI and TEMP-MED), i.e. when the disambiguating region requires *mehr* to remain unaccented (cf. the significant main effect of 'disambiguation' in Table 3). This effect seems to suggest that the readers used information in the disambiguating phrase for the assignment of accent on *mehr*, but it remains unclear what type of information this might be and what makes this information misleading. In any case, this effect shows that the disambiguating material does not have a facilitating effect on the appropriate realization of *mehr* in unprepared reading. A comparison of the accentuation patterns in unprepared reading with those of the prepared session indicates that the readers were most likely unaware of the disambiguating information in the unprepared session. Also, the manifestation of the verb stress effect in unprepared reading suggests that readers use implicit rhythmic cues more readily than whatever information they have about the disambiguating phrase when determining the accentuation status of *mehr* in this task.

The phonetic analysis of the accented and unaccented versions of *mehr* confirms the validity of the judges' perceptions. As expected, accented *mehr* is realized with a strong rise in pitch and longer duration compared to unaccented versions. The duration data at the clause break provide an indication that the accentuation involves a syntactic commitment on the part of the speakers. The region before the disambiguating clause is significantly prolonged when realizations of *mehr* are incompatible with the disambiguating region. This slowdown is indicative of a garden path effect. The data suggest that the readers/speakers in fact assign syntactic features to *mehr* according to their realization of accent on this item and experience integration difficulty if the disambiguating region does not conform to the prior prosodic realization.

Overall, the first experiment confirms that reading prosody is dependent not only on the syntactic structure and the lexically determined syllable and stress information of the words in the written string, but also on the supralexical linguistic rhythm emerging from the concatenation of single words. Specifically, the experiment presents evidence for the hypothesis that rhythmic expectancy, i.e. the avoidance of stress clashes, affects the prosodic realization and, consequently, syntactic parsing in unprepared oral reading. Beyond a general preference for the unaccented temporal reading of *mehr*, the local rhythmic environment demonstrably constrains the respective assignment of the syntactic features. That is, if the syntactic structure is underspecified, the reader chooses the accentuation and, consequently, the syntactic analysis that best conforms to syntactic and prosodic well-formedness constraints. This interpretation of the results implies that readers evaluate the syntactic structure of written material as a function of the prosodic environment which is generated by a process of phonological recoding. At first glance, this idea is at odds with existing research on reading prosody that emphasizes the dependence of prosody on the syntactic analysis (Kondo and Mazuka, 1996; Koriat et al., 2002). Those experiments on reading prosody, however, are chiefly concerned with syntactically unambiguous structures and focus on the relation of larger syntactic constituents and prosodic phrasing. More local prosodic features like stress and linguistic rhythm may therefore affect the assignment of syntactic structure in the ambiguous region without contradicting research on the relation of syntax and prosodic phrasing. As mentioned above, different prosodic features might have different repercussions at different processing stages in reading comprehension.

In any case, experiment I does not allow firm conclusions to be drawn about the precise relation of prosodic and syntactic processes in reading. The dependent measures evaluated so far are bound to speech production in oral reading, which is known to lag behind sentence comprehension (Levin and Addis, 1979). It can thus only indirectly inform us about the interplay of syntax and prosody in the comprehension processes. Moreover, while oral reading necessarily involves reading prosody, the involvement of prosody in silent reading is less evident. Data that is arguably more time sensitive and therefore more informative about the role of implicit prosody in written sentence comprehension comes from the relevant sense organ, i.e. the eye-movement record (Rayner, 1998; Clifton et al., 2007).

4. Experiment II

The notion of a speechlike phonological representation in silent reading implies that readers have rhythmic expectancies. They should especially avoid representations of adjacent stressed syllables whenever more rhythmic alternatives are accessible. The present experiment tests this hypothesis using the same material as in experiment I applying eye-tracking methodology for silent reading. The example sentences are repeated in (2).

(2) Der Polizist sagte, dass man...
The policeman said that one ...

a. ... nicht mehr NACHweisen kann, wer der Täter war. TEMP-INI
... couldn't prove anymore who the culprit was.

b. ... nicht mehr erMITteln kann, wer der Täter war. TEMP-MED
... couldn't determine anymore who the culprit was.

c. ... nicht MEHR nachweisen kann, als die Tatzeit. COMP-INI
... couldn't prove more than the date of the crime.

d. ... nicht MEHR ermitteln kann, als die Tatzeit. COMP-MED
... couldn't determine more than the date of the crime.

As in experiment I, we hypothesize that readers should choose the syntactic category of ambiguous words in such a way as to accord with rhythmic preferences. In the case of the structures in (2), *mehr* should be computed as an unaccented temporal adverbial more often when followed by a verb with initial stress to avoid a stress clash configuration. This in turn should lead to increased reading difficulties in the disambiguating region if the comparative reading of *mehr* is required. That is, reading the disambiguating clause in the clash condition COMP-INI should be associated with higher processing costs compared to reading the rhythmically alternating COMP-MED. No such difference is expected between TEMP-INI and TEMP-MED as neither of them violates rhythmic preferences. Therefore, an interaction between the factors 'disambiguation' and 'verb stress' is predicted. Beyond this interaction, the clear preference for the unaccented version of *mehr*, which was attested in experiment I, should lead to increased reading difficulties in conditions COMP-INI and COMP-MED as compared to TEMP-INI and TEMP-MED.[6]

To test the influence of the rhythmic environment on the resolution of a local syntactic ambiguity in silent reading, eye-tracking methodology was employed. This method involves monitoring readers' eye-movements as they scan written text on a screen. The difficulty of identifying and integrating a given word is strongly correlated with the fixation patterns on and around that word. Tracking these patterns with a high temporal and spatial resolution therefore allows sentence comprehension processes to be studied in real time (Rayner, 1998). Syntactic parsing difficulties in the disambiguating region should be reflected in more fixations, longer fixation durations, and a higher probability of regressions in that area (Clifton et al., 2007).

4.1. Materials and methods

4.1.1. Stimuli

The set of 24 experimental items from experiment I was used for the eye-tracking experiment as well. Again, the factors 'disambiguation' and 'verb

[6] In the unprepared oral reading experiment, more instances of *mehr* were accented in the TEMP-conditions than in the COMP-conditions; this unexpected effect, however, does not alter the predictions for the silent reading experiment; the amount of appropriate realizations of *mehr* relative to the disambiguating region was still markedly higher in the TEMP-conditions (72%) than in the COMP-conditions (20%), confirming the preference for the unaccented, temporal reading.

stress' were crossed in a 2×2 design to define the four conditions.

4.1.2. Participants

Forty-eight undergraduate students of the University of Potsdam took part in the experiment for course credit or were paid for participation. All of them reported normal or corrected-to-normal vision. None of them participated in experiment I.

4.1.3. Experimental procedure

The participants were seated in front of an IView-X eye-tracker (SensoMotoric Instruments) running at 240 Hz sampling rate and with 0.025 degree resolution. To ensure stability of the eye position, participants placed their heads in a frame with a chin rest. A camera within in the frame monitored the pupil of the participant's right eye during the entire experiment. Each sentence was presented on a single line on a 17" monitor with 1024x768 pixel resolution. Stimulus presentation and recording of the eye-movements were controlled by Presentation software. The experimental sentences were divided into four lists such that experimental sentences and conditions were counterbalanced across lists and participants saw at most one sentence from each of the 24 item sets. Each participant was assigned one of the four lists, each of which contained 24 target items together with 76 filler sentences from four unrelated experiments in pseudorandomized order. A calibration procedure preceded the experiment: participants looked at 13 fixation points that appeared in random order to allow gauging of the gaze position. This procedure was repeated after every 10–15 trials or when measurement accuracy was poor. To direct the participants' eyes to the beginning of the sentence, a fixation point was shown at the position of the leftmost character immediately before presentation of the trial. Directly upon fixation of this target, the sentence was displayed. Participants were asked to silently read the sentence and click a mouse button when finished. A forced choice comprehension question followed each trial, for example *'Hat der Polizist etwas gesagt?'* (Engl.: 'Did the policeman say something?'). Answering the question by mouse click triggered the presentation of the next item.

4.1.4. Defining the dependent measures and the regions of interest

Four eye-tracking measures that are considered standard measures in the literature on eye-tracking in sentence comprehension research (Rayner, 1998;

Clifton et al., 2007) were used as dependent variables. These are i) first-pass reading time (FPRT), i.e. the sum of all fixation durations within a region until leaving the region, given that the region was fixated at least once; ii) second pass or re-reading time (RRT), that is the summed fixation time on a given region after first pass (including zero times if the region was not refixated); iii) the total fixation time (TFT); iv) the probability of regressing out of a region during first pass (RegrP), i.e. before material to the right of the region was fixated. In addition to these standard measures, the probability of skipping a word (SKIP) during first pass and the re-reading probability (RRP), i.e. the probability of refixating a region after first pass, were calculated. All dependent measures are examined on individual words in or near the disambiguating region.

FPRT and RegrP are assumed to reflect so-called 'early' processing stages and may indicate the difficulty associated with higher level lexical processing such as integrating words with the preceding context (Clifton et al., 2007). The skipping probability (SKIP) may reflect even earlier processes since the decision to fixate or skip a word during first pass is necessarily made on a word preceding the affected target word. The 'late' measures (RRT, RRP and TFT) are generally considered to reflect more general comprehension difficulties (Clifton et al., 2007).

Unfortunately, it is far from clear how to distinguish between integration difficulty on the one hand and general comprehension difficulty on the other. The interpretation of the dependent measures and the distinction between 'early' and 'late' processing stages depend on various factors such as type of ambiguity, strength of interpretation bias, type of disambiguation (morphological, syntactic, semantic) and also on the size of the region under examination. Therefore, the precise cognitive processes responsible for a particular dependent measure remain a matter of debate in eye-movement research in reading. In general, however, longer reading times and higher regression rates are associated with higher cognitive demand, while shorter fixation times and more frequent skipping may signal relative reading ease (Clifton et al., 2007; Rayner, 1998).

In this experiment, two words (full verb and modal verb comprising 12 to 18 characters altogether) intervene between the ambiguous word and the disambiguating region. Moreover, a clause boundary precedes the disambiguating phrase. Reading difficulties are expected to show up in the disambiguating phrase that follows the modal verb. The first word of this phrase disambiguates the temporal and comparative reading: for the temporal reading, it is either a (wh-)pronoun or a complementizer that introduces a sentential complement to the preceding verb complex (cf. TEMP-INI and TEMP-MED in (2)). In the comparative reading, the standard marker *als* introduces the comparative complement of *mehr*, which is the standard of comparison (cf. COMP-INI and COMP-MED in (2)). Since the first word of the disambiguating phrase is relatively short (3 or 4 characters), measures were analyzed on both the first (Region 1) and the second word (Region 2) of the disambiguating phrase. Additionally, fixation patterns on the modal verb that precedes the disambiguating region (Region 0) were analyzed in order to check for parafoveal effects (Kennedy and Pynte, 2005; Kliegl et al., 2006). This is motivated by the fact that Region 1 was frequently skipped (see below) and skipping of a word usually correlates with longer reading times on the preceding word (Kliegl and Engbert, 2005). Finally, to gauge a possible spillover effect, the eye-movement record of the last word of the disambiguating phrase (Region 3) was examined, too.

4.1.5. Data analysis

Due to miscalibrations, data from one participant was excluded from further analysis (only 5% of the subject's fixations were recorded as fixations on words).

Question response accuracies were computed. Only those trials that were responded to correctly and in which the critical verb following *mehr* was fixated during first pass were included in the statistical analysis of the eye-tracking measures. The *em* package by Logačev and Vasishth (2006) was used to calculate the dependent measures. For the statistics on FPRT and TFT, fixations shorter than 50 ms were removed and treated as missing values. In order to adjust for the skew in the data, fixation durations were log-transformed for inferential statistics (Gelman and Hill, 2007).

Statistical analysis

The fixation durations (FPRT, RRT, TFT) were analyzed using linear mixed models; skipping probability (SKIP), re-reading probability (RRP) and regression probability (RegrP) were modeled using generalized linear mixed models (GLMM) with binomial link function (Bates and Sarkar, 2007). As in experiment I, the dependent measures were evaluated against the factors 'disambiguation' (TEMP vs. COMP) and 'verb stress' (INI vs. MED) and the respective interaction. Participants and items were included as crossed random effects. Again, contrast coding was applied as in experiment I (factor 'disambiguation': comparative=1, temporal=-1; factor 'verb stress': initial=1, medial=-1).

4.2. Results

Various reading measures in several regions of the disambiguating phrase reveal an increase of processing costs in the comparative conditions relative to the temporal conditions, with most difficulty arising in the clash condition COMP-INI. Results for each region of interest will be detailed below.

4.2.1. Response accuracy

On average, participants answered 86% of the comprehension questions correctly. A GLMM that evaluates the error rates against the experimentally controlled factors does not reveal any significant influence of the fixed factors on the distribution of the erroneous answers (effect of 'disambiguation': *z-value=-0.474, p=0.64*; effect of 'verb stress': *z-value: 1.567, p=0.12*; interaction: *z-value=1.175, p=0.24*).

4.2.2. Reading measures

The reading measures for the Regions 0 through 3 are tabulated in Table 7. The results of all regions are discussed in the following. Inferential statistics for Regions 2 and 3 are tabulated in Tables 8 and 9 respectively.

Region 0: Word preceding the disambiguating phrase

The word preceding the disambiguating phrase is a mono- or disyllabic modal verb comprising 4 to 7 characters. The clash condition COMP-INI displays the highest FPRT, RRT and TFT in this region. Apparently, the values for conditions COMP-INI and COMP-MED differ more strongly than those of TEMP-INI and TEMP-MED, suggesting an interaction between the fixed factors. Inferential statistics on these measures, however, do not yield any significant effect (all t-values are distinctly <|2|). Likewise, GLMMs on SKIP and RRP do not yield any significant effects (all z-values are distinctly <|2|, p>0.05). However, RegrP gives rise to a significant main effect for the factor 'disambiguation'. A regression was made from this word significantly more frequently when the disambiguating region required the temporal reading of *mehr* (*coeff. estimate=-0.2097, std.err.=0.1038, z-value=-2.02, p-value=0.043*). The

Table 7: Raw reading measures (means) broken down by condition and region of interest.

		Region of interest			
Measure	cond.	0	1	2	3
SKIP	TEMP-INI	0.15	0.45	0.24	0.2
	TEMP-MED	0.09	0.41	0.23	0.19
	COMP-INI	0.10	0.47	0.24	0.17
	COMP-MED	0.08	0.49	0.33	0.21
FPRT (Std.Err.) in ms	TEMP-INI	224 (8)	248 (10)	238 (10)	314 (18)
	TEMP-MED	222 (8)	233 (9)	236 (9)	305 (20)
	COMP-INI	245 (10)	229 (9)	231(9)	351(19)
	COMP-MED	236 (10)	216 (9)	240 (10)	313 (17)
RegrP	TEMP-INI	0.21	0.13	0.22	0.55
	TEMP-MED	0.25	0.10	0.28	0.54
	COMP-INI	0.15	0.07	0.22	0.68
	COMP-MED	0.17	0.07	0.23	0.55
RRP	TEMP-INI	0.33	0.21	0.33	0.19
	TEMP-MED	0.36	0.28	0.32	0.19
	COMP-INI	0.34	0.26	0.46	0.29
	COMP-MED	0.29	0.22	0.32	0.23
RRT (Std.Err.) in ms	TEMP-INI	84 (10)	68 (11)	97 (12)	38 (7)
	TEMP-MED	88 (10)	71 (10)	87 (11)	65 (13)
	COMP-INI	101 (13)	81 (13)	132 (13)	113 (16)
	COMP-MED	86 (13)	57 (10)	100 (14)	79 (16)
TFT (Std.Err.) in ms	TEMP-INI	319 (14)	341 (20)	352 (18)	364 (20)
	TEMP-MED	317 (13)	332 (17)	340 (16)	388 (25)
	COMP-INI	353 (16)	342 (21)	376 (17)	486 (28)
	COMP-MED	328 (18)	300 (15)	367 (20)	413 (28)

main effect for 'verb stress' and the interaction term remain non-significant. The higher regression probability in the TEMP-conditions is reminiscent of the clause wrap-up effect that has been reported in Rayner et al. (2000). Its implications will be discussed in the discussion section together with the results from the other regions of interest.

Region 1: 1st word of disambiguating clause
The first word of the disambiguating clause (the actual disambiguating word) is a short function word in all conditions (3–4 characters). During first pass, it was skipped on average in 46% of the trials. Considering also later fixations, it was fixated at least once in 73% of the trials altogether. To test whether skipping of this word is affected by any of the controlled factors, a GLMM was fitted (with first pass skipping as the binomial response variable) yielding no significant effects for the factors 'disambiguation' (*z-value=1.486, p=0.137*) or 'verb stress' (*z-value=0.618, p=0.536*) or for the interaction term (*z-value=-0.803, p=0.422*). The evaluation of the reading times (FPRT, RRT and TFT) against the controlled factors plus the interaction does not yield any significant effect (all t-values <|2|). Similarly, RegrP and RRP lack significant effects (with z-values <|2| and p>0.05 for all main effects and interactions). Given the high skipping probability, the reading measures on this word may be unreliable.

Region 2: 2nd word of disambiguating clause
Because of the inconclusive and likely unreliable results on the disambiguating word, reading times on the second word of the disambiguating clause were examined, too (cf. Table 7 for the means). On average, the second word was skipped in 26% of the trials. In the comparative reading (COMP-INI and COMP-MED), this word is a short function word (determiner, preposition or pronoun) in the majority of cases;[7] as for conditions TEMP-INI and TEMP-MED, the word category of this position is more varied across items. Condition COMP-MED displays the highest skipping probability. Inferential statistics reveal that the interaction between 'disambiguation' and 'verb stress' approaches significance. Further analysis demonstrates that skipping occurred significantly more frequently in COMP-MED as compared to COMP-INI (factor 'verb stress' (comparative conditions only): *coeff. estimate=-0.2532,*

[7] Two comparative items feature an adverb in this position.

std.err.=0.1136, z-value=-2.229, p=0.0258). In contrast, the difference in skipping rate between TEMP-INI and TEMP-MED is negligible. The source of the interaction effect is therefore attributable to the difference between conditions COMP-INI and COMP-MED. While FPRT and RegrP do not show any considerable differences between the four conditions, the other measures reveal that the stress clash condition COMP-INI gives rise to the highest RRP, the highest RRT, and the highest TFT of the four conditions (cf. Table 7) in this region. No systematic effects of the fixed factors were found for TFT in this region (cf. Table 8). Inferential statistics for RRP show that the interaction between the factors 'verb stress' and 'disambiguation' approaches significance. Further analysis shows that re-reading this region is significantly more likely in the clash condition COMP-INI compared to condition COMP-MED (factor 'verb stress': *coeff. estimate=0.32, std.err.=0.10, z-value=3.11, p=0.0019*), whereas the difference between TEMP-INI and TEMP-MED is not systematic. This confirms again that the disparity between COMP-INI and COMP-MED is the main source of the interaction. As for RRT, the interaction between the fixed factors closely approximates significance. Singling out the two comparative conditions COMP-INI and COMP-MED, a linear mixed model confirms that the factor 'verb stress' significantly contributes to the difference between the two conditions (*coeff. estimate=0.799, std.err.=0.267, t-value=2.99*). Again the difference between TEMP-INI and TEMP-MED is marginal.

Region 3: Last word of disambiguating clause
The last word of the sentence was examined in order to determine whether the experimental factors show effects beyond the immediate vicinity of the disambiguating word. The average skipping probability is 19%. It is again the clash condition COMP-INI that displays the highest values in all other measures under scrutiny (cf. Table 7). While the values in COMP-INI and COMP-MED differ considerably, the values of TEMP-INI and TEMP-MED are much more similar. FPRT does not show any significant effects. Inferential statistics on regression probability reveal significant main effects for the factors 'disambiguation' and 'verb stress', as well as for the interaction (cf. Table 9). Closer inspection indicates that the two main effects are largely due to the salient values of COMP-INI. Looking specifically at the difference between the two COMP-conditions, a GLMM yields a significant effect for the factor 'verb stress' (*coeff. estimate=0.3386, std.err.=0.1065, z-value=3.179, p=0.0015*). As for TFT, the interaction between disambiguation and verb

Table 8: Modeling results for 2nd word of disambiguating clause (Region 2).

Measure	Coefficient	Estimate	Std. Error	Test Statistics
SKIP	disambig.	0.13893	0.09278	z=1.497, p=0.1343
	verb stress	-0.10734	0.08080	z=-1.328, p=0.1840
	disamb × v-stress	-0.14148	0.08031	z=-1.762, p=0.0782
FPRT	disambig.	-0.00084	0.01691	t=-0.05
	verb stress	-0.00803	0.01701	t=-0.47
	disamb × v-stress	-0.01308	0.01694	t=-0.77
RegrP	disambig.	-0.08538	0.09903	z=-0.862, p=0.389
	verb stress	-0.07302	0.09783	z=-0.746, p=0.455
	disamb × v-stress	0.04291	0.09744	z=0.440, p=0.660
RRP	disambig.	0.17492	0.07782	z=2.248, p=0.0246
	verb stress	0.17800	0.07378	z=2.413, p=0.0158
	disamb × v-stress	0.13826	0.07337	z=1.884, p=0.0595
RRT	disambig.	0.20837	0.09024	t=2.309
	verb stress	0.20739	0.08370	t=2.478
	disamb × v-stress	0.14613	0.08323	t=1.756
TFT	disambig.	0.04257	0.02431	t=1.75
	verb stress	0.01851	0.02105	t=0.88
	disamb × v-stress	0.00977	0.02099	t=0.47

Table 9: Modeling results for last word of disambiguating clause (Region 3).

Measure	Coefficient	Estimate	Std. Error	Test Statistics
SKIP	disambig.	-0.03024	0.10452	z=-0.289, p=0.772
	verb stress	-0.06954	0.09162	z=-0.759, p=0.448
	disamb × v-stress	-0.13224	0.09128	z=-1.449, p=0.147
FPRT	disambig.	0.00915	0.02845	t=0.32
	verb stress	0.02394	0.02338	t=1.02
	disamb × v-stress	0.01587	0.02319	t=0.68
RegrP	disambig.	0.25837	0.10327	z=2.502, p=0.0124
	verb stress	0.19447	0.08944	z=2.174, p=0.0297
	disamb × v-stress	0.14439	0.08850	z=1.632, p=0.1028
RRP	disambig.	0.31979	0.11060	z=2.891, p=0.00383
	verb stress	0.09508	0.09305	z=1.022, p=0.30686
	disamb × v-stress	0.10872	0.09251	z=1.175, p=0.23989
RRT	disambig.	0.26367	0.08293	t=3.180
	verb stress	0.06013	0.06712	t=0.896
	disamb × v-stress	0.12737	0.06664	t=1.911
TFT	disambig.	0.04404	0.02890	t=1.52
	verb stress	0.02828	0.02228	t=1.27
	disamb × v-stress	0.04875	0.02207	t=2.21

stress is also significant. Again, focusing on the COMP-conditions, the linear model yields a significant effect of 'verb stress' (*coeff. estimate=0.0917, std.err.=0.0313, t-value=2.931*). RRT and RRP give rise to a significant main effect for the factor 'disambiguation' with higher RRTs and RRPs for COMP-conditions compared to TEMP-INI and TEMP-MED.

In summary, the eye-movement data shows significantly increased reading costs for the comparative conditions COMP-INI and COMP-MED, when compared with the TEMP-conditions. Over and above this main effect of 'disambiguation', lower skipping probabilities, longer reading times and a higher likelihood of regressions are attested for the clash condition COMP-INI compared to the rhythmically innocuous condition COMP-MED, in the absence of a similar difference between the control conditions TEMP-INI and TEMP-MED. Although no significant effects were found on the actual disambiguating word (arguably due to the high skipping rate), the predicted interaction between the controlled factors 'disambiguation' and 'verb stress' is attested on the second word (in SKIP, RRT and RRP) and continues to affect eye-movements until the end of the sentence (RegrP, RRT and TFT). Note that, with the exception of FPRT and RegrP in Region 2, the coefficients of the predicted interaction between 'disambiguation' and 'verb stress' all signal higher reading costs for COMP-INI as compared to COMP-MED, i.e. they are negative for SKIP and positive for the other measures (cf. Tables 8 and 9). Even in the absence of significant effects for some of the dependent variables, this consistency suggests that the salience of the clash condition COMP-INI is systematic.

4.3. Discussion

As hypothesized, the present results suggest that rhythmic preferences indeed affect the silent parsing of written text. The eye-movement record in the disambiguating region of the test sentences attests systematic reading costs for the comparative disambiguation, reflecting the general preference for the temporal, unaccented version of *mehr*. The reading costs for comparative *mehr* are particularly high when the critical verb following *mehr* bears initial stress (COMP-INI). The prosodic representation of the comparative reading of *mehr* requires an accent on this word. Initial stress on the immediately following verb, therefore, would force a stress clash in this condition. The increased reading times relative to the rhythmically innocuous conditions indicate that the stress manipulation is critical for the assignment of syntactic structure. Readers avoid implicit accentuation of *mehr* when this would generate a stress clash. Accordingly, the unaccented temporal analysis of *mehr* is eminently preferred in this situation, which leads to increased processing demand if the comparative reading turns out to be the correct one.

The evaluation of several eye-tracking measures at different points within the disambiguating region supports this interpretation of the results. Before reviewing the supporting evidence, we address the inconclusive results that were obtained for the actual disambiguating word. This word is a short function word (3–4 characters) that introduces either a sentential complement to the preceding verb complex (in the temporal disambiguation) or a comparative complement starting with the word *als*. The shortness together with the fact that these words invariably introduce a new clause may be the reason for the high number of missing fixations on this word. Generally, short function words are heavily susceptible to skipping (Rayner, 1998). Moreover, it has been established that readers make relatively long saccades into a new clause, thus increasing the likelihood of skipping phrase initial words (Rayner et al., 2000). Together, these factors might well explain the missing fixations on the disambiguating word. It has to be noted though that fixating a word is not a necessary condition for processing it. Especially short words with a high frequency may be sufficiently recognized in parafoveal view. The reading data on the word preceding the disambiguating region (Region 0) provides a slight indication that the disambiguating word is already processed at this position: in line with the predictions, FPRT, RRT and TFT are (non-significantly) higher in the comparative conditions with the highest values in the clash condition COMP-INI. Regression probabilities in Region 0 are higher in the conditions with temporal reading of *mehr* as compared to the COMP-conditions. The regression probabilities for conditions TEMP-INI and TEMP-MED are strikingly reminiscent of the clause wrap-up effect reported by (Rayner et al., 2000, p. 1072); such an effect might be expected given the clause break at this position. However, why this wrap-up effect does not appear in conditions COMP-INI and COMP-MED remains open to speculation. One reason might be that readers experienced difficulties associated with the disambiguating word (Region 1) in parafoveal view in the COMP-conditions, which prevented them from programming a regressive saccade.

The second word of the disambiguating clause gives rise to an interaction between 'disambiguation' and 'verb stress' in skipping probability (SKIP), with significantly less frequent skipping in COMP-INI as compared to COMP-MED. As discussed above, effects concerning SKIP may reflect relatively early sentence processing stages. In fact, the decision to skip a word during first pass must be made while fixating a preceding region. It is therefore likely that readers make this decision while processing the actual disambiguating word.[8] Also, re-reading probability and re-reading time are significantly increased in the clash condition COMP-INI. These measures are said to reflect more general comprehension difficulty (Clifton et al., 2007), suggesting that readers struggle to overcome the reading difficulties they encounter in this condition. It is also possible that these measures reflect reanalysis, which requires more effort in COMP-INI due to the stress clash. Similarly, on the

[8] Determining when exactly readers process the actual disambiguating word is not trivial given the high number of missing first pass fixations on this word and the likelihood that words are processed even in parafoveal view.

last word of the disambiguating clause, the high regression probability and the high total fixation time indicate persisting reading difficulty in the clash condition COMP-INI.

The increased reading times for the comparative versions compared to the temporal disambiguation found in Regions 2 and 3 are most likely due to the general preference for the temporal reading (as would be predicted by Bader (1996) and the unprepared oral reading experiment). However, since the lexical material between the two disambiguations is not necessarily comparable, this explanation should be taken with some caution.

In summary, the results of the silent reading experiment II appear to be compatible with the stress clash effect found in experiment I and thus confirm the involvement of supralexical, stress-based linguistic rhythm in parsing written text.

5. General Discussion

Previous research has uncovered effects of lexical stress on eye-movement patterns in silent reading. Those results were taken as an indication of an early speech-like prosodic representation of the text. The findings of the present experiments augment the evidence for a speech-like representation. They confirm that, during silent reading, readers mentally construct patterns of implicit lexical prominences that evolve from the concatenation of individual words. While processing written text, readers obey prosodic-phonological preferences such as the principle of rhythmic alternation and they especially avoid sequences of adjacent stressed syllables. These results are important in that they demonstrate the involvement of supralexical, stress-based linguistic rhythm in silent reading of ordinary text. Moreover, the findings not only attest the mere existence of the rhythmic effect but also point to the functional role it may have in written sentence comprehension. The results shed light on the interplay of syntactic and prosodic processing during written sentence comprehension. The findings indicate that the process of analyzing the lexical-syntactic features of critical words is sensitive to the local prosodic environment even in the written modality, where no explicit prosodic cues exist. As for the timing of the rhythmic effect in relation to the syntactic analysis, two competing accounts will be discussed in the following:

First, it is conceivable that, once the reader encounters the ambiguous word *mehr*, he commits himself to the preferred temporal, unaccented reading and would only reconsider the decision if forced by syntactic counterevidence in the disambiguating region. According to this view, the parser would initially disregard rhythmic preferences or prosodic cues in general (cf. Augurzky, 2006, for such a proposition). The increased reading costs for the conditions with the comparative disambiguation would come about solely due to the syntactic and concomitant prosodic reanalysis. Correspondingly, the prosodic reanalysis would force implicit accentuation of *mehr* and thus induce the rhythmically imperfect stress clash representation in condition COMP-INI, which would cause the additional increase in processing costs. While the effects in the 'late' reading measures (as, for instance re-reading time and total reading time) may be attributable to syntactic-prosodic reanalysis, the swiftness of the stress clash effect in silent reading (especially concerning skipping probability in Region 2) casts considerable doubt on the idea of the initial disregard of prosodic information. In order for the clash-effect to emerge only during reanalysis, the corresponding syntactic-prosodic revision would have to proceed very fast, i.e. probably immediately on encountering the disambiguating word. In contrast to this interpretation, syntactic-prosodic revisions are generally said to be rather costly and time-consuming (Bader, 1998). Also, this account cannot easily explain the results of experiment I. In the unprepared oral reading task, the choice of the accent status of *mehr* was clearly influenced by the immediate rhythmic environment, while, at this stage, disambiguating information further downstream was not systematically taken into account. Moreover, word prosodic information such as stress (the decisive factor for the evaluation of stress clash) is most likely computed rapidly online in silent reading (Ashby and Clifton, 2005; Ashby and Martin, 2008). Therefore, readers have the necessary information for evaluating potential stress clashes as soon as the critical words are combined within the reader's processing window, i.e. arguably before the evaluation of the disambiguating material in the present experiment.

The alternative account proposes that lexical stress contributes to parsing decisions more immediately. Hence, the eye-tracking effects may, but need not necessarily reflect prosodic reanalysis; instead, the early effects within the disambiguating region may be interpreted as reflecting integration difficulty. Accordingly, the interpretation of *mehr* would be affected already prior to the disambiguating region on the basis of the following verb's stress pattern. On this view, the parser is not always fully committed to the preferred reading when encountering the ambiguous word. Rather, the parser makes a variable choice (with a bias to the temporal reading), which depends on various sources of information (cf. Van Gompel et al., 2001, and references therein). If – in spite of the general preference for the temporal, unaccented version – the parser happened to have initially foregrounded the comparative, accented analysis of *mehr*, it sets this analysis back when a stress clash would ensue due to the implicit accentuation of *mehr* and initial stress on the following verb.[9] Subsequently, the reader would expect a continuation compatible with the unaccented, temporal interpretation. The parser would therefore face integration difficulties when the disambiguating region forces the comparative interpretation of *mehr*, hence the increased processing costs in the clash condition COMP-INI. Significantly smaller processing costs ensue in the comparative condition with medial stress on the verb COMP-MED, as there is no rhythmic trigger that would demand unaccented *mehr* in addition to the general preference for the temporal analysis.

The very swiftness of the rhythmic effect in the eye tracking data makes this latter account more appealing. It is compatible with both the results from the unprepared oral reading experiment and the eye-tracking data (silent reading); moreover, it also acknowledges the finding that lexical-prosodic features are computed very early in reading (Ashby and Clifton, 2005; Ashby and Martin, 2008).

This interpretation of the results is consistent with the *Implicit Prosody Hypothesis* (IPH) by Fodor (2002), which predicts on-line effects of implicit prosody on the syntactic analysis. According to the IPH, the parser computes the most natural (default) prosody in line with the incremental analysis; the accruing prosody may affect syntactic decisions, biasing the parser to a syntactic analysis compatible with the prosodic representation. Previous experiments demonstrating on-line effects of implicit prosody on parsing (e.g. Hirose, 2003; Hwang and Steinhauer, 2011) were concerned with the interdependence of implicit prosodic phrasing and the attachment of larger syntactic constituents in reading. The prosodic trigger in these studies, phrase length, requires at least low-level syntactic processing, namely the merging of words for the formation of phrases, whose length can be evaluated. In the present case, the default prosody (i.e. the preference for rhythmic alternation) was shown to affect more elementary building blocks of the syntactic structure, namely the assignment of the syntactic category of ambiguous words.

[9] It is also conceivable that the verbal stress pattern already impinges on the initial processing of *mehr*, if it is assumed that the access to the syntactic features of *mehr* overlaps with access to prosodic-phonological features of the following verb. The reading data within the disambiguating region, however, are not informative on this matter.

Beyond the resolution of the lexical ambiguity, the avoidance of accentuation of the critical word *mehr* in the face of a potential stress clash impinges – at least indirectly – on the syntactic predicate-argument structure. Note that the syntax-phonology interface in German requires arguments of verbs to be accented unless they are pronouns or given in the discourse (e.g. Truckenbrodt, 2006). Correspondingly, it is unlikely that readers interpreted *mehr* as a comparative filling the object position of the transitive verb when they did not accent it; conversely, accent on *mehr* might guide the parser to posit an argument.[10]

On the basis of these considerations, we suggest that prosodic-phonological and syntactic processing (and possibly semantic processing) are coupled and may alternately lead the way in written sentence comprehension. In the case of the temporarily ambiguous sentences of the present experiments, phonological well-formedness conditions like the principle of rhythmic alternation are considered for the computation of syntactic structure and may affect very early stages of the analysis – in this particular case: the determination of lexical-syntactic features.

This notion of written sentence comprehension is in line with models of sentence comprehension holding that different types of information (syntactic, semantic, phonological, etc.) may exert their influence on sentence comprehension as soon as they become available in the input (e.g. MacDonald et al., 1994; McRae et al., 1998; Van Gompel et al., 2001). The present results demonstrate the influence of supralexical preferences concerning linguistic rhythm and accordingly call for the implementation of these effects in models of written sentence comprehension.

Finally, the results may also be taken as evidence for a more integrated account of sentence comprehension and production in reading. At the outset, it was predicted – based on speech production research – that (implicitly) producing a stress clash is generally avoided whenever more rhythmic alternatives are accessible. Both experiments confirm this view. Moreover, the results indicate that rhythmic preferences have repercussions for the com-

[10]Whether the stress clash directly affects syntactic computation, or whether it does so only via lexical disambiguation, cannot be determined on the basis of the present data. In any case, there is reason to assume that the domain of the stress clash effect is restricted to the phonological phrase (Hayes, 1989; Nespor and Vogel, 2007). Accordingly, stress clashes may possibly affect a syntactic constituent that corresponds to a phonological phrase (e.g. the constituent comprising a verb and its object) but not of larger syntactic clauses.

prehension process. Readers expect the written text, or more precisely the implicit phonological representation thereof, to be rhythmic; this apparently has consequences for the syntactic analysis. Thus, sentence comprehension in reading is at least in this respect driven by constraints that are standardly understood as being chiefly relevant to speech production.

We hope to have revealed the need to study the workings of linguistic rhythm beyond the acoustics of speech. Linguistic rhythm (and phonology in general) may be deemed an integral part of any linguistic behavior irrespective of the modality of perception and performance.

6. Acknowledgements

The research reported here was conducted within the "Prosody in Parsing" project at Potsdam University, funded by the German Research Foundation (DFG). I thank Shravan Vasishth for generously providing his eye-tracking lab and appreciate the training in statistical data analysis I received from him. Pavel Logačev's help concerning the management of the eye tracking data is gratefully acknowledged. Kirsten Brock copy-edited the manuscript. Markus Bader, Caroline Féry, Fatima Hamlaoui, Frank Kügler, Sara Myrberg, Fabian Schubö, and Shravan Vasishth provided helpful comments. The manuscript has greatly benefited from many careful and knowledgable comments from three anonymous reviewers.

7. References

Anttila, A., Adams, M., Speriosu, M., 2010. The role of prosody in the English dative alternation. Language and Cognitive Processes 25 (7-9), 946–981.

Ashby, J., Clifton, C., 2005. The prosodic property of lexical stress affects eye movements during silent reading. Cognition 96 (3), B89–B100.

Ashby, J., Martin, A., 2008. Prosodic phonological representations early in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance 34 (1), 224–236.

Ashby, J., Rayner, K., 2004. Representing syllable information during silent reading: Evidence from eye movements. Language and Cognitive Processes 19 (3), 391–426.

Augurzky, P., 2006. Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing. Ph.D. thesis, MPI for Human Cognitive and Brain Sciences, Leipzig.

Bader, M., 1996. Prosodic effects and the distinction between primary and non-primary phrases. In: Proceedings of the Conference on Architectures and Mechanisms of Language Processing (AMLaP). University of Torino.

Bader, M., 1998. Prosodic influences on reading syntactically ambiguous sentences. In: Fodor, J., Ferreira, F. (Eds.), Reanalysis in sentence processing. Dordrecht: Kluwer, pp. 1–46.

Bates, D., Sarkar, D., 2007. lme4: Linear mixed-effects models using S4 classes. R package version 0.9975-11.

Beckman, M., Pierrehumbert, J., 1986. Intonational structure in Japanese and English. Phonology Yearbook 3, 255–309.

Bergmann, A., Armstrong, M., Maday, K., 2008. Relative clause attachment in English and Spanish: A production study. In: Proceedings of the 4th conference on Speech Prosody. Campinas, Brazil, pp. 505–508.

Breen, M., Clifton, C., 2011. Stress matters: Effects of anticipated lexical stress on silent reading. Journal of Memory and Language 64 (2), 153–170.

Chafe, W., 1988. Punctuation and the prosody of written language. Written Communication 5 (4), 395–426.

Clifton, C., Staub, A., Rayner, K., 2007. Eye movements in reading words and sentences. In: Gompel, R. V., Fisher, M., Murray, W., Hill, R. L. (Eds.), Eye movements: A window on mind and brain. Elsevier, Amsterdam, Ch. 15, pp. 341–372.

Dilley, L., McAuley, J., 2008. Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language 59 (3), 294–311.

Fodor, J., 1998. Learning to parse? Journal of Psycholinguistic Research 27 (2), 285–319.

Fodor, J. D., 2002. Psycholinguistics cannot escape prosody. In: Proceedings of the 1st International Conference on Speech Prosody. Aix-en-Provence, pp. 83–88.

Forster, K., Forster, J., 2003. DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments and Computers 35 (1), 116–124.

Fuchs, L., Fuchs, D., Hosp, M., Jenkins, J., 2001. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading 5 (3), 239–256.

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge.

Gussenhoven, C., 1983. Focus, mode and the nucleus. Journal of Linguistics 19 (2), 377–417.

Hayes, B., 1989. The prosodic hierarchy in meter. In: Kiparsky, P., Youmans, G. (Eds.), Rhythm and Meter. Academic Press, Orlando, pp. 201–260.

Hayes, B., 1995. Metrical stress theory: Principles and case studies. University of Chicago Press.

Hirose, Y., 2003. Recycling prosodic boundaries. Journal of Psycholinguistic Research 32 (2), 167–195.

Hwang, H., Schafer, A., 2009. Constituent length affects prosody and processing for a dative NP ambiguity in Korean. Journal of Psycholinguistic Research 38 (2), 151–175.

Hwang, H., Steinhauer, K., 2011. Phrase length matters: The interplay between implicit prosody and syntax in korean garden path sentences. Journal of Cognitive Neuroscience Early Access, 1–21.

Jun, S., 2003. Prosodic phrasing and attachment preferences. Journal of Psycholinguistic Research 32 (2), 219–249.

Jun, S., 2010. The implicit prosody hypothesis and overt prosody in English. Language and Cognitive Processes 25 (7), 1201–1233.

Kelly, M., Bock, J., 1988. Stress in time. Journal of Experimental Psychology: Human Perception and Performance 14 (3), 389–403.

Kennedy, A., Pynte, J., 2005. Parafoveal-on-foveal effects in normal reading. Vision research 45 (2), 153–168.

Kliegl, R., Engbert, R., 2005. Fixation durations before word skipping in reading. Psychonomic Bulletin & Review 12 (1), 132–138.

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, 12–35.

Kondo, T., Mazuka, R., 1996. Prosodic planning while reading aloud: On-line examination of Japanese sentences. Journal of Psycholinguistic Research 25 (2), 357–381.

Koriat, A., Greenberg, S., Kreiner, H., 2002. The extraction of structure during reading: Evidence from reading prosody. Memory and Cognition 30 (2), 270–280.

Levin, H., Addis, A., 1979. The eye-voice span. MIT Press, Cambridge.

Logačev, P., Vasishth, S., 2006. The em package for computing eyetracking measures. University of Potsdam, Potsdam, Germany.

MacDonald, M. C., Perlmutter, N. J., Seidenberg, M. S., 1994. Lexical nature of syntactic ambiguity resolution. Psychological Review 101(4), 676–703.

McRae, K., Spivey-Knowlton, M. J., Tanenhaus, M. K., 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language 38, 283–312.

Nespor, M., Vogel, I., 2007. Prosodic phonology. Mouton de Gruyter, Berlin.

Niebuhr, O., 2009. F0-based rhythm effects on the perception of local syllable prominence. Phonetica 66 (1–2), 95–112.

Quené, H., Van den Bergh, H., 2004. On multi-level modeling of data from repeated measures designs: A tutorial. Speech Communication 43 (1), 103–122.

Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124 (3), 372–422.

Rayner, K., Kambe, G., Duffy, S., 2000. The effect of clause wrap-up on eye movements during reading. The Quarterly Journal of Experimental Psychology Section A 53 (4), 1061–1080.

Schlüter, J., 2005. Rhythmic grammar: the influence of rhythm on grammatical variation and change in English. Walter de Gruyter.

Schmidt-Kassow, M., Kotz, S., 2009. Attention and perceptual regularity in speech. NeuroReport 20 (18), 1643–1647.

Selkirk, E., 1984. Phonology and syntax. MIT Press Cambridge, Mass.

Selkirk, E., 1995. Sentence prosody: Intonation, stress, and phrasing. In: Goldsmith, J. (Ed.), The handbook of phonological theory. Blackwell, Oxford, pp. 550–569.

Speyer, A., 2010. Topicalization and stress clash avoidance in the history of English. Mouton De Gruyter, Berlin.

Steinhauer, K., 2003. Electrophysiological correlates of prosody and punctuation. Brain and Language 86 (1), 142–164.

Steinhauer, K., Friederici, A., 2001. Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. Journal of Psycholinguistic Research 30 (3), 267–295.

Truckenbrodt, H., 2006. Phrasal stress. In: Brown, K. (Ed.), The encyclopedia of languages and linguistics, 2nd Edition. Vol. 9. Elsevier, Amsterdam, pp. 572–579.

van Casteren, M., Davis, M., 2006. Mix, a program for pseudorandomization. Behavior Research Methods 38 (4), 584–589.

Van Gompel, R., Pickering, M., Traxler, M., 2001. Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. Journal of Memory and Language 45 (2), 225–258.

Warren, P., Grabe, E., Nolan, F., 1995. Prosody, phonology and parsing in closure ambiguities. Language and Cognitive Processes 10 (5), 457–486.

Wiese, R., 2000. The phonology of German. Oxford University Press.