

Implicit prosody and contextual bias in silent reading

Kate McCurdy
University of Potsdam, Germany

Gerrit Kentner
University of Frankfurt, Germany

Shravan Vasishth
University of Potsdam, Germany

Eye-movement research on implicit prosody has found effects of lexical stress on syntactic ambiguity resolution, suggesting that metrical well-formedness constraints interact with syntactic category assignment. Building on these findings, the present eyetracking study investigates whether contextual bias can modulate the effects of metrical structure on syntactic ambiguity resolution in silent reading. Contextual bias and potential stress-clash in the ambiguous region were crossed in a 2×2 design. Participants read biased context sentences followed by temporarily ambiguous test sentences. In the three-word ambiguous region, main effects of lexical stress were dominant, while early effects of context were absent. Potential stress clash yielded a significant increase in first-pass regressions and re-reading probability across the three words. In the disambiguating region, the disambiguating word itself showed increased processing difficulty (lower skipping and increased re-reading probability) when the disambiguation engendered a stress clash configuration, while the word immediately following showed main effects of context in those same measures. Taken together, effects of lexical stress upon eye movements were swift and pervasive across first-pass and second-pass measures, while effects of context were relatively delayed. These results indicate a strong role for implicit meter in guiding parsing, one that appears insensitive to higher-level constraints. Our findings are problematic for two classes of models, the two-stage garden-path model and the constraint-based competition-integration model, but can be explained by a variation on the two-stage model, the unrestricted race model.

Keywords: silent prosody; stress-clash; implicit meter; context; reanalysis; garden-path model; competition-integration model; unrestricted race model; re-reading probability; skipping rate.

Introduction

When a reader encounters a temporary syntactic ambiguity, which factors influence their parsing decisions, and at what stage of the syntactic analysis? Assuming, as many theories do, that a multitude of information

sources affect the parser's choice, how do these factors interact in guiding the reader's understanding? Answering this question has important implications for theories of sentence comprehension, such as constraint satisfaction accounts (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey & Tanenhaus, 1998) and classical reanalysis models (e.g., Frazier & Rayner, 1982; Traxler, Pickering, & Clifton, 1998). We examine an important test case where global information, made available by preceding context, either contradicts or supports strictly local information stemming from the lexical-prosodic structure that is implicit in the written string.

We thank Felix Engelmann, Eva Saur, Titus von der Malsburg, Umesh Patil, and Paul Metzner for extensive assistance in the study's design and execution; Titus von der Malsburg also provided very helpful comments. The research reported here is based on the master's thesis of the first author (EMCL program). The experiment was funded partly by the University of Potsdam (vasishth lab) and partly by the Deutsche Forschungsgemeinschaft, through the project Prosody in Parsing (2009-2012), which was part of the DFG-Schwerpunktprogramm 1234: Sprachliche Kompetenz: Zwischen Grammatik, Signalverarbeitung, und neuronaler Aktivität. We are extremely grateful to Manfred Krifka, Director of the Zentrum für Allgemeine Sprachwissenschaft, for allowing us to use his laboratory in Berlin to conduct the study.

The effects of global discourse context on online parsing are well-established. Some of the earliest findings investigate the role of one- versus multiple-referent contexts in facilitating processing of relative clauses (Crain & Steedman, 1985; Altmann & Steedman, 1988). While the bulk of the literature in this area has continued to concentrate on referential context and noun modification (e.g., Britt, 1994; Spivey & Tanenhaus, 1998; Binder, Duffy, & Rayner, 2001; Snedeker

& Trueswell, 2004), various contextual manipulations have also been found to affect reader expectations for a wide range of grammatical phenomena, including sentential coordination (Hoeks, Vonk, & Schriefers, 2002), non-canonical word order (Kaiser & Trueswell, 2004), temporal adverbial attachment (Altmann, Nice, Garnham, & Henstra, 1998), and the syntactic category of noun-verb homographs (Boland & Blodgett, 2001). Interestingly, however, Van Gompel and Pickering (2007) have pointed out that, although discourse effects have been reported at the earliest stages of processing, strong preferences from other information sources (e.g., lexical biases) can generally override contextual biases. This interesting opposition between the role of discourse context versus local biases motivated our investigation.

One important local constraint is the role of prosody in interpretation during silent reading. In spoken language comprehension, prosody has been found to have rapid effects on parsing syntactic ambiguity, with a time scale on the order of lexical bias (Kjelgaard & Speer, 1999; Steinhauer, Alter, Friederici, et al., 1999). The investigation of prosodic constraints in silent reading has been largely premised upon the possibility that phonological representations during reading are richly featured and fundamentally “speech-like” (Chafe, 1988). Fodor (1998) proposed that implicit prosody may have a role in syntactic parsing preferences during silent reading. This proposal has found some support in studies that examined effects of implicit prosodic phrasing on attachment ambiguities (Hirose, 2003; Augurzky, 2006). Regarding prosodic effects on the lexical level, Ashby and Clifton (2005) found that words with two stressed syllables induced longer gaze durations than words of equal length with one stressed syllable. The authors interpret these results as suggesting that metrical stress assignment reflects the completion stage of lexical access. Breen and Clifton (2011) find evidence of an additional processing cost concerning syntactic ambiguity resolution, stemming from metrical reanalysis, i.e., online reassignment of stress patterns. Apparently, these effects even precede signs of syntactic reanalysis for noun-verb ambiguous homographs (*record*, *record*).

Stress clash in German: The case of ‘nicht mehr’

Recently, Kentner (2012) has shown that the syntactic categorization of an ambiguous lexical item can be influenced by implicit patterns of word stress. His experiment used eye-tracking to investigate participants’ online interpretation of the syntactic category of the word *mehr* (‘more’) immediately following the word *nicht* (‘not’). When immediately preceded by *nicht*, there are two possible readings for *mehr*. Crucially, these senses are differentiated in spoken language by accent placement, while in written text it is possible for the syntactic category of *mehr* to remain ambiguous until the

complement of the main verb is determined.

In the preferred interpretation, *mehr* is unstressed, therefore unaccented, and *nicht* and *mehr* together form a temporal adverbial lexeme with the meaning ‘not anymore’. This sense is illustrated in the following example item from Kentner’s stimuli (stressed syllables are underlined):

- (1) Der Polizist sagte, dass man nicht mehr ermitteln kann, wer der Täter war.
 ‘The policeman said that one couldn’t determine anymore who the culprit was.’

In this sentence, *nicht mehr* modifies the main verb *ermitteln*, which receives the main phrase accent. The sentential argument following the comma forms the complement to the main verb. Bader (1996) found a general preference for this temporal reading, with unaccented *mehr*, in a self-paced reading study, and this preference was replicated in the ratings given to materials in Kentner’s study (2012: Experiment 1).

In the alternative interpretation, *mehr* itself is the syntactic complement to the main verb, and as such receives the emphasis associated with the main phrase accent. *Nicht mehr* in this construction corresponds to the English phrase ‘not more’, and *mehr* requires a following comparative complement beginning with the word *als* (‘than’).

- (2) ...dass man nicht mehr ermitteln kann, als die Tatzeit.
 ‘...that one couldn’t determine more than the date of the crime.’

Kentner hypothesized that, even in silent reading, a speechlike preference for alternating stressed and unstressed syllables would lead readers to avoid constructing implicit prosodic representations which contain a stress clash, i.e. a pair of stressed syllables directly next to each other (Kelly & Bock, 1988). In order to test this hypothesis, the conditions represented in (1) and (2) were compared to otherwise identical sentences in which the main verb was replaced with a semantically plausible alternative verb that had stress on the first syllable instead of the second:

- (3) ...dass man nicht mehr nachweisen kann, wer der Täter war.
 ‘...that one couldn’t prove anymore who the culprit was.’
- (4) ...dass man nicht mehr nachweisen kann, als die Tatzeit.
 ‘...that one couldn’t prove more than the date of the crime.’

In the condition shown in (4), the syntactic category of *mehr* (complement to the main verb) requires it to be accented, and this leads to a clash with the initial syllable of the verb, which necessarily carries lexical

stress. If readers construct an implicit prosodic representation which is indeed speechlike and subject to the constraints of spoken prosody, then the presence of a stressed syllable directly following *mehr* should lead them to prefer the temporal interpretation shown in (3), in which *mehr* is unaccented. Kentner's findings were consistent with this account: upon reaching the disambiguating postverbal region, readers showed greater processing difficulty in condition (4) compared to condition (3). This asymmetry was not found when comparing (1) and (2), in which no stress clash occurred. The increased likelihood of garden-pathing in (4) was reflected in lower first-pass skipping probabilities, longer re-reading times, and higher probability of regression during the first pass through the region. These results suggest that implicit prosody can affect syntactic parsing during the stage of syntactic category assignment.

Motivation for the present study

Kentner's findings raise further questions regarding the role of implicit prosody in higher levels of sentence comprehension. While the existence of implicit prosodic representation at the metrical level has been clearly supported by recent research (Ashby & Clifton, 2005; Breen & Clifton, 2011), and Kentner's results demonstrate its potential to influence sentential comprehension, little is known about the possible relationships between implicit lexical stress and other factors which have been shown to affect reading. In particular, it is not clear whether implicit metrical structure is automatically generated as a rule for all words in all circumstances, or whether it can be 'switched off' by other, global factors. A very basic observation that speaks to this possibility is the fact that people read faster than they speak: skilled readers of English typically read at an average rate of 300 words per minute (e.g., Rayner, 1975; Carver, 1992), while recent corpus analysis reveals average speaking rates of 190-215 words per minute in conversation (Yuan, Liberman, & Cieri, 2006). In order to accommodate the increased speed of comprehension during reading, readers may construct implicit metrical representations which are partial or somehow underspecified, or they may assign metrical structure for some words and not others; the precise mechanics of implicit prosody have yet to be explored in current research. In fact, even if implicit meter is fully speechlike and obligatorily computed under all circumstances, evidence from spoken prosody suggests that its influence may be mediated by higher-level constraints: studies which have found a metrical effect upon word order in the production of English note that this effect occurs only in the absence of semantic animacy (McDonald, Bock, Kelly, et al., 1993; Shih, Grafmiller, Futrell, & Bresnan, 2009). This potential flexibility in assignment of implicit metrical structure raises the possibility of influence from global factors

such as discourse context, which has been shown to rapidly affect expectations in online sentence processing. If implicit metrical structure is sensitive to higher-level constraints, then a facilitative discourse context may mitigate the prosodic garden-path observed for Kentner's stimuli. Conversely, if local implicit prosody is generated automatically without reference to global contextual biases, then a prosodically-induced garden path can occur regardless of the influence of context. Implicit prosody thus offers a window into the interacting roles of global and local biases in online comprehension.

Experiment

This experiment seeks to probe the interaction between implicit prosody and previous discourse context in online sentence processing, using the familiar technique of temporarily ambiguous garden-path sentences. As discussed above, Kentner (2012) demonstrated that the implicit prosodic context surrounding an ambiguous phrase (*nicht mehr*) can guide syntactic expectations. The current study builds on those findings by investigating whether this effect of implicit prosody is modulated by global discourse context.

Design & Materials

The stimuli from Kentner (2012) were modified and expanded upon for the current experiment. Thirty-six target sentences were constructed with the temporary ambiguity illustrated in (1) and (2), using the same semantically-related, stress-alternating verb pairs as in Kentner's experiment. For each target sentence, two separate preceding sentences were constructed. These sentences provided a discourse context for the target sentences, and were intended to produce a global bias toward one of the two possible senses of *nicht mehr*. Sentences engendering a contextual bias toward the comparative reading contained a comparison with *als* (although never with *mehr*). Contextual bias for the temporal reading was created by emphasizing temporal dimensions of the events described (e.g., with phrases such as *In der Vergangenheit* 'In the past', or with temporal adverbs such as *ständig* 'constantly' or *häufig* 'frequently'). Adding the context manipulation to Kentner's original design yields 2 (discourse context) $\times 2$ (verb stress) $\times 2$ (disambiguation) = 8 separate conditions for each of the 36 items. An example item is given in (5) below.¹

¹ It may be argued that our context sentences are more closely related to priming than to more conventional context manipulations which target, for example, number of potential referents. Indeed, this is likely to be true. In the 'consistent' context condition, priming may occur on several levels: the presence of *als* may produce lexical priming; the presence of *als* in the relevant, comparative sense, along with the comparative morpheme, may produce morphosemantic priming;

(5) Context

COMPARATIVE:

Der Sportmanager setzt Peer unter Druck, länger zu trainieren, als alle anderen Athleten.

The sports manager puts Peer under pressure to train longer than all the other athletes.

TEMPORAL:

Peer's Sportmanager hat leider schon oft zu viel von Peer gefordert.

Peer's sports manager has often asked too much of Peer.

Target

Peer denkt, dass der Trainer...

Peer thinks that the trainer...

- a. ... nicht mehr erlauben sollte, dass er täglich dreimal trainiert.
...should not allow anymore that he trains three times a day.
- b. ... nicht mehr erlauben sollte, als tägliches Fitnesstraining.
...should not allow more than daily fitness training.
- c. ... nicht mehr zulassen sollte, dass er täglich dreimal trainiert.
...should not permit anymore that he trains three times a day.
- d. ... nicht mehr zulassen sollte, als tägliches Fitnesstraining.
...should not permit more than daily fitness training.

For the purposes of our experiment, however, all eight conditions are not necessary; only target sentences that disambiguate to the comparative reading are relevant to the current investigation. We used 18 sets of items as target sentences; in addition, in order to not bias the reader towards the comparative reading; we included a further set of 18 items that disambiguated to the temporal reading; these served as fillers.

With the experimental design clarified, two further issues arise: (i) What criterion should be used to determine which items become test sentences and which become fillers? (ii) How can we verify that our context manipulation is effective, i.e., genuinely biases reader interpretation toward the intended meaning? On the latter question the literature provides little guidance — to the best of our knowledge, no previous studies attempt to use discourse context to create the expectation of a comparative construction. In order to address this issue, we conducted a preliminary norming study to independently assess the effectiveness of our context sentences and use the results to divide the stimuli: those items which displayed a higher sensitivity to the context manipulation would become test items, with the less sensitive serving as fillers.

Validation of Materials. A sentence rating study was conducted to investigate the efficacy of the context ma-

nipulation. Context sentences from one of three separate conditions (temporal bias, comparative bias, and a third exploratory condition with implied contrast sets — this condition was disregarded in the main experiment, and results from this condition are excluded from the current analysis) were paired with target sentences from one of two conditions (temporal disambiguation, comparative disambiguation). In order to assess the influence of context independently of any prosodic bias, only target sentences with medial stress on the critical verb were presented.

The context-norming experiment was conducted over the internet, using the experimental software WebExp (Keller, Gunasekharan, Mayo, & Corley, 2009). Three separate lists were created with a 3×3 Latin Square design. Forty native German speakers took part in the sentence rating experiment. Participants were randomly assigned to one of the nine lists. Target sentences were assessed on a Likert scale of 1–7, with 1 indicating *leicht, gut verständlich* ('good, easy to understand') and 7 *schwer, unverständlich* ('difficult to understand'), reflecting the grading system in German schools, in which smaller numbers correspond with better scores. Participants were explicitly asked to evaluate the ease of understanding each target sentence with respect to its preceding context sentence. Items were presented in two sequential stages. First the context sentence appeared onscreen, and participants were instructed read the sentence, then press any key to advance. After the keypress, the context sentence stayed onscreen while the target sentence appeared below it, along with a visual number line representing a scale from 1 *leicht* to 7 *schwer* to eliminate potential confusion concerning the task. Participants then read the target sentence and rated its understandability by pressing a number key, at which point the next item was presented. As compensation for the experiment, participants were entered in a random drawing to win a gift certificate.

The results from the rating experiment indicate that participants' judgments of the acceptability of the target sentence were affected both by the context manipulation and by an overall preference for the temporal reading of *nicht mehr*. A linear mixed model was fit to the ratings data, with contextual consistency (i.e., comparative context with comparative disambiguation, and respectively for temporal) and target sentence disambiguation (temporal or comparative) as fixed effects and participant and item as random ef-

and the presence of *als* and its comparative complement at the end of the sentence (which is the case in 17 of the 18 context sentences) may produce morphosyntactic priming. Nonetheless, the main goal of our context manipulation is a constraint which affects reader expectations at a global level, in contrast to the lower-level prosodic constraint. We are not aware of any theoretical framework for structural priming, in contrast to discourse context, which would yield predictions substantially different from those discussed below.

Table 1
Results of the linear mixed model analysis of the ratings data.
See text for details.

	Coef.	SE	t	p
Contextual Consistency	-0.12	0.05	-2.2	<.05
Disambiguation	-0.17	0.05	-3.1	<.01
Interaction	-0.02	0.06	-0.3	0.7

fects. Contrast coding was applied as follows: Contextual consistency was coded as 1 if the context and target senses matched, and -1 if they did not match. Target sentence disambiguation was coded 1 for the preferred temporal disambiguation and -1 for comparative disambiguation. The model results are summarized in Table 1; also see Figure 1. Contextual consistency yielded a significant main effect upon sentence ratings, indicating that target sentences were systematically judged to be better (in this rating system, scored lower) when the preceding context was consistent with the sense of *nicht mehr* that ultimately prevailed in the disambiguation. There was also a reliable preference for the temporal reading of *nicht mehr* as opposed to the comparative sense, a result which is consistent with earlier findings (Bader, 1996; Kentner, 2012).

Eyetracking Stimuli. The results of the rating study were used to determine which items would be test items in the eyetracking experiment. For each item, a ‘match advantage’ was calculated by subtracting the average target sentence rating when the context was consistent with the disambiguation from the average rating for cases in which the context and the disambiguation were inconsistent. The resulting number showed the average benefit in ratings that could be expected for a particular item when the contextual bias and the target sentence disambiguation were consistent. Items were then ranked according to their match advantage. The median match advantage score was 0.31 (minimum -1.7, maximum 1.8, mean 0.12). Eighteen items with match advantage scores above the median were classified as test items, due to their higher sensitivity to the context manipulation. The eighteen items with match advantage scores below the median were designated as fillers.

Four lists were created for the experiment. Each list contained eighteen test items and eighteen fillers, as determined by the sentence rating study — these categories remained constant, such that the same eighteen items were test items in each list. For test items, the target sentence always disambiguated to the comparative reading of *nicht mehr*, while in filler items it always disambiguated to the temporal reading, so that participants encountered both readings with equal frequency during the experiment.

The two experimental manipulations, contextual bias and verb stress, were distributed across the four lists in a crossed 2×2 design. The four resulting conditions are designated as follows: C.1, comparative contextual bias with initial verb stress; C.2, comparative contextual bias and medial verb stress; T.1, temporal context, initial verb stress; T.2, temporal context, medial verb stress. To avoid introducing any confounds, conditions were fully balanced across filler items as well. This resulted in four lists of thirty-six target sentences of which (a) eighteen were preceded by a context sentence biased toward the comparative interpretation of *nicht mehr*, and the other eighteen were preceded by a context sentence biased toward the temporal interpretation; (b) eighteen sentences had stress on the first syllable of the critical verb, and the other eighteen had stress on the second syllable; and (c) eighteen sentences disambiguated to the comparative reading of *nicht mehr* (the test items), and the other eighteen disambiguated to the temporal reading. As eighteen (the number of test items per list) is not evenly divisible by four (the number of conditions in the 2×2 design), it was not possible to fully balance the experimental conditions within in any one list, but conditions were crossed to the closest approximation and fully counterbalanced across the four lists. An example test item and example filler item are presented in (6) and (7) respectively. All items and fillers are available from the first author.

(6) Example Test Item

Context

COMPARATIVE (CONSISTENT WITH DISAMBIGUATION):

Die täglichen Hausaufgaben sind sogar noch schwieriger als gedacht.

The daily homework assignments are even harder than expected.

TEMPORAL (INCONSISTENT WITH DISAMBIGUATION):

Tim macht nur manchmal seine Hausaufgaben.

Tim only does his homework from time to time.

Target

Tim meint, dass man...

Tim believes one...

1. Initial-stress:

...den Lehrern nicht mehr anbieten sollte als das Erledigen eines Teils der Hausaufgaben.

...shouldn't offer the teacher more than completion of part of the homework.

2. Medial-stress:

...den Lehrern nicht mehr versprechen sollte als das Erledigen eines Teils der Hausaufgaben.

...shouldn't promise the teacher more than completion of part of the homework.

(7) Example Filler Item

Context

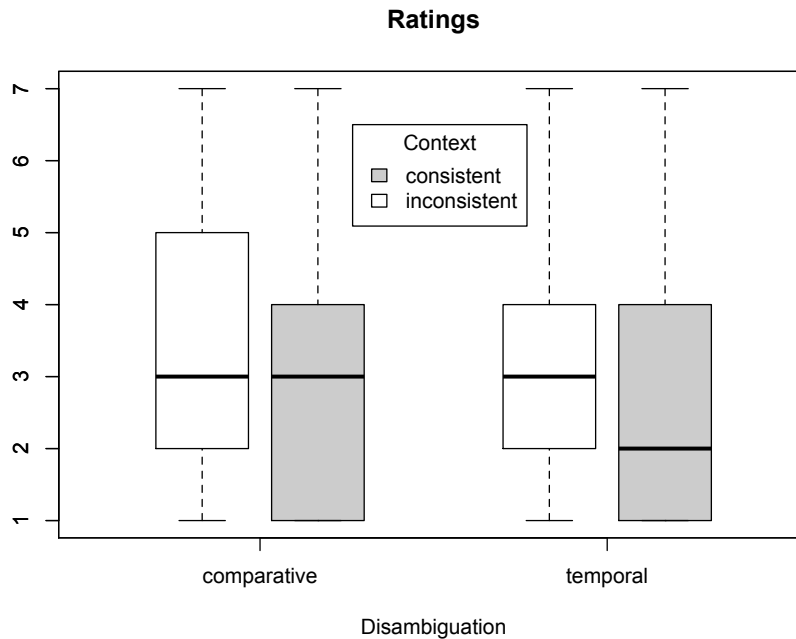


Figure 1. Boxplots showing the distributions of the ratings for the four conditions.

COMPARATIVE (INCONSISTENT WITH DISAMBIGUATION):

Paul vermutet, dass seine Mieter am Wochenende die Musik noch lauter drehen als unter der Woche.

Paul thinks that his tenants play their music louder on the weekend than during the week.

TEMPORAL (CONSISTENT WITH DISAMBIGUATION):

Paul vermutet, dass seine Mieter zu oft spät abends laute Musik hören.

Paul thinks that his tenants listen to loud music too often late at night.

Target

Paul meint, dass man als Vermieter...

Paul thinks that one, as a landlord, ...

1. Initial-stress:

...nicht mehr hinnehmen sollte, dass die Mieter ständig laute Musik hören.

...shouldn't accept anymore that the tenants constantly listen to loud music.

2. Medial-stress:

...nicht mehr gestatten sollte, dass die Mieter ständig laute Musik hören.

...shouldn't allow anymore that the tenants constantly listen to loud music.

Participants. Fifty-two native speakers of German from the Berlin area took part in the experiment. Participants were compensated either with course credit (if they were University of Potsdam students) or with cash. All participants reported normal or corrected-to-

normal vision. Visual inspection of scanpath records subsequent to data collection, but prior to all statistical analysis, revealed consistently inadequate calibrations for four participants. Data from these participants was then excluded from analysis, so that forty-eight participants ultimately contributed to the final analysis.

Procedure. Participants were randomly assigned to one of the four lists and seated in front of an IView-X eye-tracker (Senso-Motoric Instruments), using a chin rest to ensure stability. Data on the position of the participant's right eye was recorded at a sampling rate of 240 Hz (0.025 degree tracking resolution, and <0.5 degree gaze position accuracy.). Participants were seated 55 cm from a 17" color monitor which had a resolution of 1024 × 768. The angle per character was 0.3 degrees (3.8 characters per degree of visual angle).

Stimulus presentation and synchronization with eye-movement recordings were controlled by a separate computer running Presentation software. Calibration was carried out at the beginning of the experiment, and calibration quality was visually monitored throughout the course of the experiment, with recalibration every ten trials, or more frequently if necessary. Participants were given five practice trials to establish familiarity with the task before the experiment began.

At the start of each trial, the participant was required to fixate upon a black dot in the center of the left side of the screen to ensure calibration quality. Upon successful fixation, the context sentence appeared on the screen, at which point the participant read it through and pressed the continuation button. The fixation point

appeared once more at the same location, and after one second the point was replaced by the target sentence. The participant was required to answer a yes-no comprehension question after each item. While context sentences were occasionally broken into two lines, target sentences always appeared on one line. Items were presented in a randomized order and interspersed with forty-eight filler items from several unrelated experiments. The experiment generally took 45 minutes or less to complete.

Predictions

According to two-stage or reanalysis models (e.g., Frazier & Rayner, 1982), the parser responds to ambiguity with an initial structural commitment and processing difficulty occurs only when later information forces reanalysis of the initially assigned structure. In contrast, constraint-satisfaction models assume that multiple parses are activated in parallel and processing difficulty stems from competition between parses, usually due to activation from conflicting sources of information. We consider predictions from Frazier and colleagues' garden-path model (Frazier & Rayner, 1982) and the constraint-based competition-integration model (Spivey & Tanenhaus, 1998; McRae, Spivey-Knowlton, & Tanenhaus, 1998). Throughout, we follow the assumptions of the eye-tracking literature that early and late dependent measures are indicative of early and late parsing events (Sturt, 2003; Clifton, Staub, & Rayner, 2007; Vasishth, von der Malsburg, & Engelmann, 2012).

The garden-path model (Frazier & Rayner, 1982) posits that the parser's initial commitment is guided by purely syntactic factors.² In this experiment, the principle of Late Closure would support attaching *mehr* to the last item analyzed, i.e., *nicht*, yielding the temporal adverbial phrase *nicht mehr*. This is consistent with the general preference for the temporal interpretation (see Bader, 1996; Kentner, 2012; see also the norming data presented above). If the parser always takes the temporal adverbial analysis initially, irrespective of the presence or absence of initial stress on the following verb, in the ambiguous region no effects of prosody are predicted. At the point of disambiguation, reanalysis should be triggered in all conditions, as all test items resolve to the dispreferred comparative reading. According to the garden-path model, multiple sources of information contribute at a later stage towards guiding reanalysis (Van Gompel & Pickering, 2007); this predicts a main effect of context for the disambiguating region, but in late dependent measures.

An alternative set of predictions under consideration is based on an instantiation of constraint-satisfaction theories, the competition-integration model (Spivey & Tanenhaus, 1998; McRae et al., 1998). This account views processing difficulty as an indication of competition between syntactic parses activated in paral-

lel. The interesting prediction of this model is that greater difficulty should occur when information from two different sources conflicts. In the ambiguous region (the region constituting *nicht mehr*, the main verb, and the modal verb immediately before *als*), condition C.1 is predicted to show the largest slowdown and/or disruption in processing, because contextual information activates *nicht mehr*'s comparative reading while the prosodic preference for stress-class avoidance activates the temporal reading.³ The key prediction of the constraint-based model in the ambiguous region is competition-related early effects of context. As initial verb stress is consistent with the globally preferred temporal reading of *nicht mehr*, no main effect of prosody is predicted.

At the point of disambiguation, the constraint-based model's predictions are unclear and depend on various additional assumptions.⁴ As the constraint model's

² Construal theory (Frazier & Clifton, 1995) modifies the original garden-path theory by restricting its scope to primary syntactic relations. In this experiment, construal theory would predict that the *comparative* reading of *nicht mehr* is the preferred analysis, for this renders *mehr* an obligatory complement of the main verb (Bader, 1996). As this prediction is incompatible with the demonstrated preference for the temporal reading (Bader, 1996; Kentner, 2012; cf. the rating experiment), the construal theory approach is excluded from our analysis on empirical grounds. Note, of course, that construal theory in general is certainly not invalidated by the facts about *nicht mehr*.

³ The model may also predict that C.2 is harder than T.1 and T.2 because both C.1 and C.2, having a comparative bias, conflict with the global temporal preference. So there might be a main effect in which a general slowdown is observed for C conditions relative to T conditions. However, a lack of knowledge as to the relative biases of all relevant constraints makes the model difficult to specify. We therefore ignore this possible prediction.

⁴ In the disambiguation region, the syntactic and semantic evidence for resolution toward the comparative reading contributes an additional source of information to the constraint model. Here condition T.1 is predicted to show the most processing difficulty, as both contextual and prosodic information favor the temporal reading, in conflict with the comparative disambiguation; condition C.2 should be the easiest, as both context and prosody are consistent with the comparative reading. Predictions for the other two conditions hinge upon assumptions regarding the relative strength of the prosodic and contextual biases. If an early influence of context is successful in modulating the prosodic bias, then an interaction should be observed, with context facilitating processing significantly more in initial-stress verb conditions — C.1 should be much easier with respect to T.1 than C.2 relative to T.2. If the relative strength of influence goes the other direction and prosody is the stronger constraint, then the interaction should go in the other direction: C.1 and T.1 should both show processing difficulty, and the relative ease of C.2 compared to T.2 should be greater than that of C.1 compared to T.1. If both constraints are equally strong, then main effects of context and prosody may be present with no interaction.

predictions for the disambiguating region encompass a wide range of possible outcomes and are hence relatively difficult to falsify, the key test for this model lies in the ambiguous region, where it predicts early effects of context and the absence of a prosodic main effect.

To summarize, the garden-path model predicts (i) no early effect of prosody or context in any of the critical regions, and (ii) a late effect of context (temporal context more costly than comparative). The constraint-based model makes clear predictions only for the ambiguous region: greater difficulty in C.1 compared to other conditions (early effect of context), and no effect of prosody.

Results

Four measures are presented: (i) first-pass fixation probability (FFP), the probability of fixating on a word during initial read-through (i.e. the probability of *not* skipping a word); (ii) first-pass reading time (FPRT), the summed duration of fixations on a particular word from the initial fixation until the first point at which the eyes exit the region, either to the right or to the left (only non-zero FPRTs are considered); (iii) first-pass regression probability (RegrP), the probability of regressing out of a region during initial read-through; and (iv) re-reading probability (RRP), the probability of re-reading a word. Generally, the first three measures — FFP, FPRT, and RegrP — are thought to reflect early processing stages, such as lexical access, while re-reading probability (RRP) is associated with later stages of processing, specifically, post-lexical processes. Although no clear linkage has yet been established between individual dependent measures and specific cognitive events in the eye-tracking record (Clifton et al., 2007; Boland, 2004; Vasishth et al., 2012), we follow the widely adopted convention of associating these measures with early vs late processes. We do not present the other dependent measures commonly used in eye-tracking research because they did not show any statistically significant effects (i.e., no information would be gained from the presentation of further dependent measures).

Standard eye-tracking measures were computed in **R** using the `em` package (Logačev & Vasishth, 2006). Dependent measures were calculated for each word individually from the ambiguous region through the point of disambiguation. In each target sentence, this entailed the following sequence: the ambiguous word *mehr*; the main verb, site of the critical stress manipulation; the modal verb; the disambiguating word *als*; and the short function word following the disambiguating word. For first-pass reading time, all fixations shorter than 50 ms were discarded from analysis. All statistical analyses were performed in **R** (R Development Core Team, 2012). Linear mixed effects models were computed with the `lme4` package (Bates & Sarkar, 2007). The reading time measures were log-transformed to

achieve approximately normal residuals, and generalized linear mixed models with a binomial link function were fit for binary response variables. Three fixed factors were specified, with the following contrast coding: verb (1 for initial stress, -1 for medial stress), context (1 for inconsistent with the comparative disambiguation, -1 for consistent), and the interaction of these two factors. Participant and item intercepts were included as random factors in all models. Means and standard errors for dependent measures across all regions are shown in Figure 2.

Average accuracy for comprehension questions was 84%, indicating that participants' attention remained engaged during the course of the experiment. A generalized linear mixed effects model with a binomial link function revealed no significant trends in response accuracy related to our manipulation.

Ambiguous region: mehr. *Mehr* is the site of the experiment's critical attachment ambiguity. Model results are shown in Table 2. A main effect of verb stress is found in probability of first-pass regressions and re-reading probability, suggesting that the stress information of the main verb is being accessed parafoveally and affecting processing on *mehr*.⁵ This interpretation is consistent with Ashby and Martin's (2008) finding of parafoveal access to syllabic structure, and Breen and Clifton's (2011) finding of parafoveally-triggered metrical effects directly on the ambiguous word in their Experiment 2. No interactions with context are present, and no effects appear on first-pass reading time or first fixation probability.

Ambiguous region: main verb. As on the previous word, a main effect of verb stress appears on re-reading probability at the verb: the critical verb is more likely to be re-read if it has initial stress. No main effects or interactions are present on first-pass regressions or first fixation probability. Model results are shown in Table 3.

First-pass reading times show a significant interaction and no reliable main effects. To clarify the nature of this interaction, two models were contrast-coded with a nested effects structure: one compared the effect of context within prosodic conditions (C.1 vs. T.1, C.2 vs. T.2), while the other compared the effect of verb stress within context conditions (C.1 vs. C.2, T.1 vs. T.2). The model output can be found in Table 4.

These post-hoc nested comparisons reveal that the slowest reading times for this region occur in condition

⁵ A reviewer suggests that these effects may simply due to uncontrolled differences in verb frequency and/or length; word form frequencies were obtained from the Leipzig Wortschatz corpus (<http://wortschatz.uni-leipzig.de>). However, the significant effect of verb stress remains unaffected in both first-pass regressions and re-reading probability even if we add centered log verb frequency and verb length (character length) as predictors.

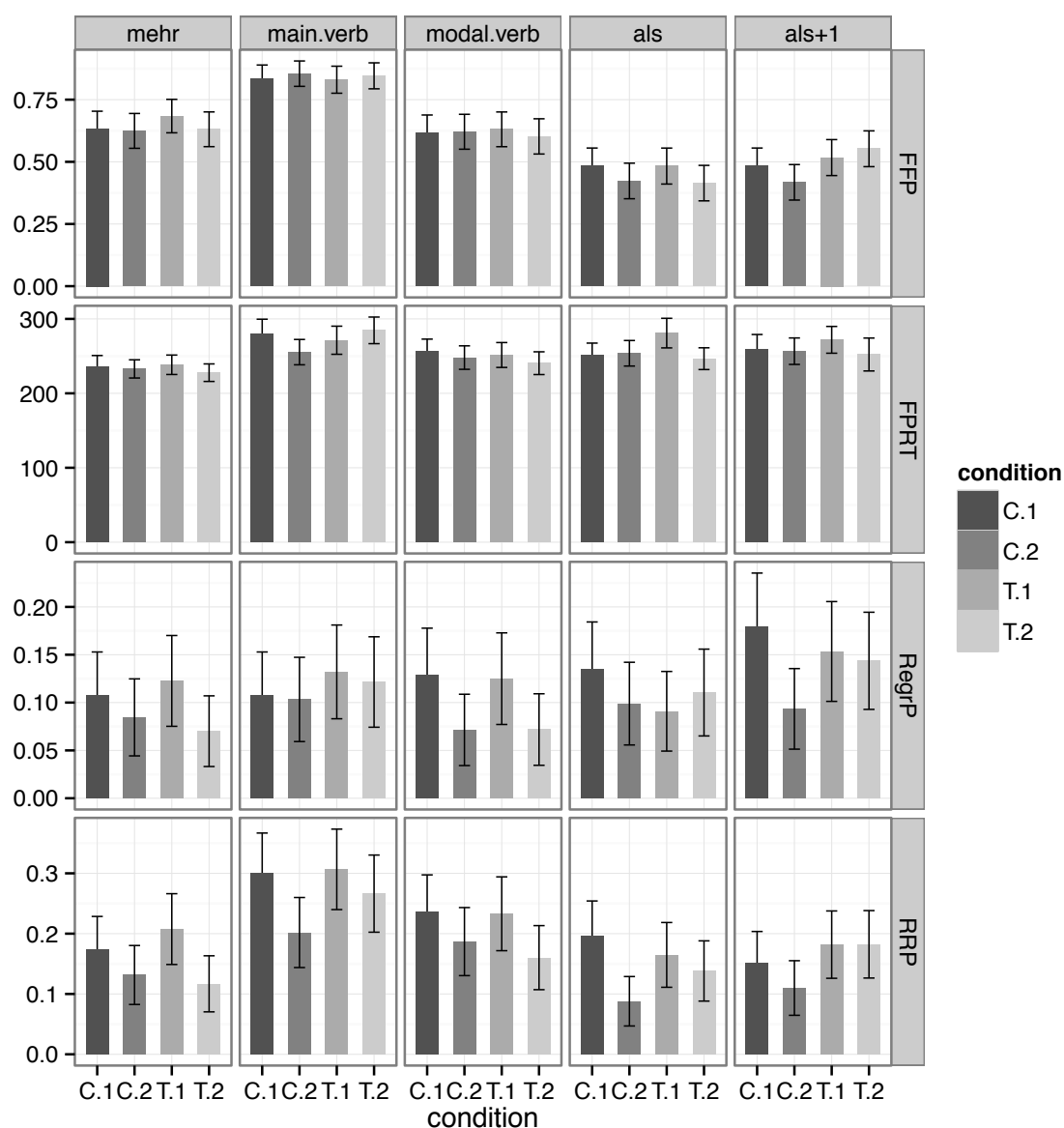


Figure 2. Means and standard errors for the relevant dependent measures, by region and condition. FFP stands for first-fixation probability (i.e., the probability of not skipping); FPRT for first-pass reading time; RegrP for regression probability; and RRP for re-reading probability.

T.2, while the fastest occur in condition C.2. The effect of prosody on first-pass reading time is significant when preceded by a comparative-biasing context (reliably slower reading times for C.1 compared to C.2), but not when the preceding context is temporal (no reliable T.1–T.2 difference). To state this differently, the effect of context is significant when verb stress is medial (reliably slower reading times for T.2 compared to C.2), but not when verb stress is initial (no reliable C.1 vs. T.1 difference).

Modal verb. For this region and all subsequent regions through disambiguation, sentences in which the

critical verb was not fixated on the initial pass were excluded from analysis. This criterion removes approximately 15% of trials from consideration.

A significant main effect of verb stress appears in two of the four dependent measures: initial stress on the verb results in a higher probability of first-pass regressions out of the region, as well as a higher probability of re-reading. First-pass fixation probability and first-pass reading time show no effects related to our manipulation, and no effects of discourse context appear either as a main effect or an interaction on any of the dependent measures. Model results are shown in

Table 2
Summary of model results for *mehr*

<i>mehr</i>	Coef.	SE	t or z	p
FFP				
Context	0.07	0.07	1.0	0.3
Verb Stress	0.07	0.07	1.0	0.3
Interaction	0.05	0.07	0.7	0.5
FPRT				
Context	0.00	0.01	-0.1	0.9
Verb Stress	0.01	0.01	0.7	0.5
Interaction	0.01	0.01	1.0	0.3
RegrP				
Context	-0.02	0.13	-0.1	0.9
Verb Stress	0.26	0.13	2.0	<.05
Interaction	0.12	0.13	0.9	0.4
RRP				
Context	0.02	0.10	0.2	0.8
Verb Stress	0.30	0.10	2.9	<.01
Interaction	0.10	0.10	1.0	0.3

Table 3
Summary of model results for main verb

main verb	Coef.	SE	t or z	p
FFP				
Context	-0.03	0.10	-0.3	0.8
Verb Stress	-0.07	0.10	-0.7	0.5
Interaction	0.00	0.10	0.0	1
FPRT				
Context	0.02	0.01	1.4	0.2
Verb Stress	0.01	0.01	0.6	0.5
Interaction	-0.03	0.01	-2.2	<.05
RegrP				
Context	0.11	0.11	0.9	0.3
Verb Stress	0.04	0.11	0.3	0.7
Interaction	0.01	0.11	0.1	0.9
RRP				
Context	0.12	0.08	1.5	0.1
Verb Stress	0.22	0.08	2.6	<.01
Interaction	-0.10	0.08	-1.2	0.2

Table 5.

Disambiguating region: *als*. As all of our test items disambiguate to the comparative sense of *nicht mehr*, the first word of the disambiguating region is always *als* ('than').

A marginal effect of verb stress on first fixation probability suggests that *als* is more likely to be fixated during first-pass reading when stress falls on the initial syl-

Table 4
Summary of post-hoc nested model results for FPRT at the main verb

FPRT at main verb	Coef.	SE	t	p
Nested context				
C.1-T.1	-0.01	0.02	-0.5	0.6
C.2-T.2	0.05	0.02	2.5	<.05
Verb Stress	0.01	0.01	0.6	0.5
Nested verb				
Context	0.02	0.01	1.4	0.2
C.1-C.2	0.04	0.02	2.0	<.05
T.1-T.2	-0.02	0.02	-1.1	0.3

Table 5
Summary of model results for modal verb

modal verb	Coef.	SE	t or z	p
FFP				
Context	0.01	0.08	0.1	0.9
Verb Stress	0.02	0.08	0.3	0.8
Interaction	0.05	0.08	0.6	0.5
FPRT				
Context	-0.02	0.02	-0.9	0.4
Verb Stress	0.01	0.02	0.8	0.4
Interaction	-0.01	0.02	-0.5	0.6
RegrP				
Context	-0.01	0.13	-0.1	0.9
Verb Stress	0.34	0.14	2.5	<.05
Interaction	-0.01	0.13	-0.1	1
RRP				
Context	-0.03	0.10	-0.3	0.8
Verb Stress	0.21	0.10	2.1	<.05
Interaction	0.06	0.10	0.6	0.6

lable of the main verb, and more likely to be skipped when main verb stress is medial. A significant main effect of verb stress in the same direction appears in re-reading probability — participants are more likely to revisit *als* when main verb stress is initial. Regression probability shows no statistically reliable trends. Model results are shown in Table 6.

Two marginally significant interactions are observed for first-pass reading time and re-reading probability. Post-hoc nested comparisons (Table 7) reveal that these interactions reflect different underlying patterns in the results for these two dependent measures. First-pass reading time is characterized by slower reading on *als* in condition T.1 relative to all other conditions, in which reading time is effectively identical. Conversely, the interaction on re-reading probability reflects a modulation of the main effect of verb. When preceded by

Table 6
Summary of model results for *als*

<i>als</i>	Coef.	SE	t or z	p
FFP				
Context	-0.01	0.08	-0.1	0.9
Verb Stress	0.14	0.08	1.8	<.1
Interaction	0.01	0.08	0.1	0.9
FPRT				
Context	0.01	0.02	0.6	0.6
Verb Stress	0.03	0.02	1.4	0.2
Interaction	0.03	0.02	1.7	<.1
RegrP				
Context	-0.08	0.13	-0.6	0.5
Verb Stress	0.03	0.13	0.3	0.8
Interaction	-0.15	0.13	-1.2	0.2
RRP				
Context	0.07	0.11	0.7	0.5
Verb Stress	0.30	0.11	2.6	<.01
Interaction	-0.19	0.11	-1.7	<.1

Table 7
Summary of post-hoc nested comparisons for *als*

	Coef.	SE	t or z	p
FPRT				
C.1-T.1	0.04	0.03	1.6	0.1
C.2-T.2	-0.02	0.03	-0.8	0.4
Verb Stress	0.03	0.02	1.4	0.2
Context				
C.1-C.2	0.01	0.02	0.6	0.6
T.1-T.2	0.06	0.03	2.2	<.05
RRP				
C.1-T.1	-0.12	0.14	-0.8	0.4
C.2-T.2	0.26	0.17	1.5	0.1
Verb Stress	0.30	0.11	2.6	<.01
Context				
C.1-C.2	0.49	0.17	2.9	<.01
T.1-T.2	0.10	0.15	0.7	0.5

a context consistent with the disambiguation, medial stress is associated with lowest rates of re-reading (C.2), while initial stress yields the highest probability of re-reading (C.1); this prosodic effect is not as drastic when the preceding context has a temporal bias and is thus inconsistent with the disambiguation (T.1 and T.2). No main effects of context are present for any measure.

Disambiguating region: als+1. The second word of the disambiguating region directly follows *als*, and is hence

Table 8
Summary of model results for *als+1*

<i>als+1</i>	Coef.	SE	t	p
FFP				
Context	0.19	0.08	2.4	<.05
Verb Stress	0.04	0.08	0.5	0.6
Interaction	-0.11	0.08	-1.4	0.2
FPRT				
Context	0.01	0.02	0.3	0.7
Verb Stress	0.01	0.02	0.6	0.6
Interaction	0.03	0.02	1.6	0.1
RegrP				
Context	0.09	0.11	0.8	0.4
Verb Stress	0.22	0.11	1.9	<.1
Interaction	-0.17	0.11	-1.6	0.1
RRP				
Context	0.23	0.11	2.1	<.05
Verb Stress	0.11	0.11	1.0	0.3
Interaction	-0.10	0.11	-0.9	0.4

designated *als+1*. In all cases with the exception of one, this word is a determiner or personal pronoun comprising 3–4 characters; in the one exceptional item, it is the 7-character modal discourse particle *sowieso*. Model results are shown in Table 9.

In this region, the influence of context becomes clearly visible. Significant main effects of context appear on first fixation probability and re-reading probability. While no reliable verb effects or interactions are found in the primary models for these measures, the influence of verb stress persists in the form of a near-significant main effect of prosody upon regression probability. Post-hoc nested models (Table 10) for first fixation probability and re-reading probability offer additional evidence for a continuing role of prosody — though neither contrast yields a significant interaction, nested comparisons reveal that the main effect of context is driven by verb-medial conditions (greater advantage for C.2 over T.2 compared to C.1 and T.1) for both first fixation and re-reading probability.

Discussion

We first summarize the results. The ambiguous region is characterized predominantly by effects of verb stress. Initial stress on the main verb is associated with higher probability of regression on the first pass for *mehr* and the modal verb, and higher probability of re-reading for *mehr*, the main verb, and the modal verb. The influence of context is limited to an interaction on first-pass reading times at the main verb. In the disambiguating region, initial stress on the main verb is associated with higher rates of fixation and regression

Table 9
Summary of model results for als+1

als+1	Coef.	SE	t or z	p
FFP				
Context	0.19	0.08	2.4	<.05
Verb Stress	0.04	0.08	0.5	0.6
Interaction	-0.11	0.08	-1.4	0.2
FPRT				
Context	0.01	0.02	0.3	0.7
Verb Stress	0.01	0.02	0.6	0.6
Interaction	0.03	0.02	1.6	0.1
RegrP				
Context	0.09	0.11	0.8	0.4
Verb Stress	0.22	0.11	1.9	<.1
Interaction	-0.17	0.11	-1.6	0.1
RRP				
Context	0.23	0.11	2.1	<.05
Verb Stress	0.11	0.11	1.0	0.3
Interaction	-0.10	0.11	-0.9	0.4

Table 10
Summary of post-hoc nested comparisons for als+1

	Coef.	SE	z	p
FFP				
C.1-T.1	0.08	0.11	0.7	0.5
C.2-T.2	0.29	0.11	2.7	<.01
Verb Stress	0.04	0.08	0.5	0.6
Context				
C.1-C.2	0.19	0.08	2.4	<.05
T.1-T.2	0.15	0.11	1.3	0.2
T.1-T.2	-0.07	0.11	-0.6	0.5
RRP				
C.1-T.1	0.13	0.15	0.8	0.4
C.2-T.2	0.34	0.16	2.1	<.05
Verb Stress	0.11	0.11	1.0	0.3
Context				
C.1-C.2	0.23	0.11	2.1	<.05
T.1-T.2	0.21	0.17	1.3	0.2
T.1-T.2	0.00	0.15	0.0	1

during the first pass and re-reading during the second pass, consistent with the findings of Kentner (2012). Lower probability of initial fixation and re-reading are observed when the preceding discourse context is semantically consistent with the eventual meaning of *nicht mehr* after disambiguation. Both factors, context and verb stress, show main effects across various reading measures. Taken as a whole, these results point toward an influence of both prosody and context on processing during silent reading, but at different stages.

The original aim of this experiment was to determine whether the influence of a global variable, discourse context, could modulate the effects of a local variable, lexical-level implicit prosody, during online sentence processing. The results reported here show that the effects of implicit prosody are early and pervasive, but we were unable to find any effect due to the contextual manipulation. This result extends the conclusions of (Kentner, 2012) and Breen and Clifton (2011) that metrical structure can play a role in guiding structure building.

One theoretical account which can be readily ruled out in light of these results is the garden-path model. Traditional garden-path theory predicts that the temporal reading of *nicht mehr* should be initially adopted in all cases; as the temporal reading is consistent with initial stress on the following verb, no prosodic effects are predicted. The sustained effects of verb stress found in our data clearly contradict this account.

The constraint-satisfaction model which originally motivated the present study finds limited support in the form of an interaction on first-pass reading times at the point of disambiguation; however, the bulk of the data is inconsistent with this model's predictions. The central prediction of the constraint model is an early, competition-driven effect of context at the ambiguous region, here *mehr*, with spillover effects on the main verb likely. As noted in the section outlining the predictions, this could be realized either as a main effect of context (contextual bias for the comparative reading conflicts with the global preference for the temporal reading: C.1, C.2 > T.1, T.2) or as an interaction (contextual bias for the comparative reading conflicts with prosodic bias for the temporal reading: C.1 > C.2, T.1, T.2), or both; crucially, comparative contextual bias is predicted to lead to longer reading times, as it competes with temporal bias both at the global level of pre-existing lexical and syntactic preference, and at the local level of prosody when the main verb carries initial stress. Notably, contextual bias supporting the dispreferred comparative reading is also expected to lead to slowed reading times on *mehr* due to the well-established subordinate bias effect upon lexical access (Rayner, Pacht, & Duffy, 1994).

The only region that reflects any influence of context prior to disambiguation is the main verb, which shows a significant interaction between context and verb stress for first-pass reading times. This interaction, however, goes in a different direction from that predicted by the constraint model: post-hoc comparisons reveal that significantly longer reading times in condition T.2 relative to C.2 drive the result, while C.1 and T.1 show no reliable differences with any other condition. A disadvantage for contextual bias that supports the globally preferred temporal reading, alongside an advantage for contextual bias toward the dispreferred comparative reading, is completely unexpected under the constraint model. Moreover, the early effects of

verb stress in regression probability on *mehr* and the modal verb are inexplicable in terms of competition. As the prosodic temporal bias produced by initial verb stress is wholly consistent with the global preference for the temporal reading, competition should be reduced, if anything, in the presence of initial verb stress; instead, regression probability appears to reflect increased processing difficulty in exactly those conditions.

The sole piece of evidence for the constraint-based model's predictions regarding the influence of context is the statistically marginal interaction on first-pass reading times at *als*, the point of disambiguation. *Als* is read on average 30 ms slower in condition T.1 compared to the other three conditions. This result is consistent with an interpretation in which the earlier contextual bias in condition C.1 successfully activated the comparative reading despite a prosodic bias toward the temporal reading, leading to reduced competition at *als* relative to condition T.1, in which both contextual and prosodic biases toward the temporal reading conflict with the ultimate comparative disambiguation. Nevertheless, this account is undermined not only on the grounds of statistical uncertainty, but also on evidence from second-pass measures of processing. A main effect of verb stress on re-reading probability indicates that *als* is significantly more likely to be re-read in initial verb stress conditions, and an additional marginal interaction reflects the fact that condition C.1 actually shows the highest rate of re-reading. Another issue for the interaction on first-pass reading times at *als* is the unexpected absence of a difference in reading times between T.2 and C.2 — if the influence of context is strong enough to modulate prosodic bias in initial verb stress conditions, its effects should appear in the absence of prosodic bias as well. Overall, the data is difficult to reconcile with any analysis in which a comparative contextual bias manages to switch off the prosodic garden path.

Thus, the main conclusion from our work is that implicit meter plays a strong role in guiding parsing, and this effect of implicit meter is insensitive to higher-level constraints. In the remainder of this discussion, we outline a post-hoc explanation for the results.

A post-hoc explanation in terms of the unrestricted race model

Is there any model which captures the observed pattern of results? We think a case can be made for a variation on the two-stage model, namely the unrestricted race model (Traxler et al., 1998; Van Gompel, Pickering, & Traxler, 2001). Our account is post-hoc, and relies upon a couple of additional assumptions, justified below. Nonetheless, it strikes us as the most parsimonious way to describe these results within an existing theoretical framework.

The unrestricted race model is a modified two-stage

reanalysis model in which the initial parsing process is probabilistic and unrestricted: the parser draws on all available sources of information in constructing the initial parse, and the analysis assigned to a given structure may vary based on a range of constraints. Encountering an ambiguity triggers a race in which both potential grammatical structures are built up simultaneously. As soon as one of the processes finishes, the resulting parse is adopted as the parser moves forward in the sentence. Because the process is non-deterministic, while the preferred parse is generally constructed the fastest and therefore wins the race, the dispreferred parse is adopted on some non-zero proportion of trials. For this reason, ambiguous structures are considered as easy or easier to process (Logačev & Vasishth, 2013) than those which are disambiguated, as ambiguous structures are consistent with any grammatically possible analysis the parser may have initially chosen. In other words, the unrestricted race model posits a penalty for disambiguation.

The first assumption necessary to support our application of the unrestricted race framework to these data is the conjecture that the parser treats implicit prosodic information as if it were disambiguating — that is, upon reaching the initially-stressed syllable of the main verb, stress clash avoidance leads the parser to respond as if *nicht mehr* has been disambiguated to the temporal reading. This claim may appear rather unconventional, as it relies upon the premise that non-syntactic information can induce the parser to reanalyze its existing representation as ungrammatical. This reasoning is not, however, quite as outlandish as it may first appear. For one thing, the extra-grammatical factor of semantic plausibility has been successfully employed to resolve syntactic ambiguity in a variety of studies (e.g., Rayner, Carlson, & Frazier, 1983; Ni, Crain, & Shankweiler, 1996; Van Gompel et al., 2001). An additional consideration for word-level prosody is that, in spoken language, it is unequivocally disambiguating. In the prepared reading task of Kentner's Experiment 1 (2012), speakers were given time to read the sentence silently before reading it aloud, so that the ultimate meaning of *nicht mehr* was known to them before they began to speak. When followed by a comparative disambiguation, *mehr* was accented in 90% of trials; with a temporal disambiguation, *mehr* was accented in less than 10% of trials. These findings indicate that the spoken prosody of *nicht mehr* is a highly reliable cue to its meaning and structure. Furthermore, in Kentner's unprepared reading task, when speakers were not aware of *nicht mehr's* final meaning, they were significantly less likely to produce an accent on *mehr* when it was immediately followed by a verb with initial stress. This illustrates that the well-established principle of stress clash avoidance in production (Kelly & Bock, 1988) also guides speakers' preferences with respect to *nicht mehr*. This speaker preference for alternating stressed and unstressed syllables has also been found to affect

spoken language perception: listeners interpret acoustically stress-ambiguous syllable strings in accordance with a rhythmic preference for stress clash avoidance (Dilley & McAuley, 2008; Niebuhr, 2009). Taken together, it appears that 1) the presence or absence of stress on *mehr* in the phrase *nicht mehr* is a highly reliable indicator of grammatical structure; 2) speakers reliably avoid stress clash on *nicht mehr*, in accordance with the general preference in production; and 3) listeners reliably incorporate the preference for stress clash avoidance into their interpretation of ambiguous syllable sequences. The combination of these related pieces of evidence suggests that, insofar as implicit prosody is speechlike, it isn't so unreasonable to imagine the parser treating the presence of a stressed syllable directly following *mehr*, with its implied stress clash, as the silent yet effective analog of a broadly valid acoustic cue to syntactic structure.⁶

Medial verb stress, on the other hand, does not provide such a strong cue. It is more or less compatible with both prosodic realizations and thus both parses. Although prosodic theory stipulates that the counterpart to stress clash known as lapse — two adjacent unstressed syllables — is also dispreferred in production (Lieberman & Prince, 1977), the preference for lapse avoidance is much weaker than that of clash avoidance (Nespor, Vogel, et al., 1989). This is supported by Kentner's results (2012). In Experiment 2, throughout the disambiguating region, significant effects of verb stress on various dependent measures (skipping, re-reading time, and total fixation time) appeared when the disambiguation was comparative, but not when it was temporal. This indicates that, while initial verb stress produces a temporal bias that clashes when the ultimate reading is comparative, medial verb stress exerts no countervailing bias toward the comparative reading; it is simply neutral. The data from Kentner's Experiment 1 illustrate this point as well. A model evaluating the likelihood of producing the appropriate prosody for *mehr* given the ultimate disambiguation — i.e. accenting *mehr* when the final reading was comparative, and leaving *mehr* unstressed when it was temporal — found a main effect of disambiguation and a significant interaction of disambiguation and verb stress. The main effect reflected an overall tendency to avoid accenting *mehr*, yielding more inappropriate prosodic realizations in comparative-disambiguation conditions. The interaction revealed that initial verb stress increased the rate of inappropriate de-accenting relative to medial verb stress. The absence of a main effect of verb shows that appropriate accent production in temporal-disambiguation conditions was unaffected by verb stress. In both comprehension and production data from this study, prosodic clash avoidance exerted a significant effect when verb stress was initial, while there is scant evidence that lapse avoidance played any role in the medial verb stress conditions.

The implication of this claim is that our experiment's

verb stress manipulation not only alters the implicit rhythm of the sentence at the critical point of syntactic ambiguity, it actually creates an additional disambiguating region in those conditions where stress clash is anticipated. In this interpretation, when the ambiguous word *mehr* is encountered, the race to construct a suitable parse begins — only to be abruptly halted when the stressed syllable immediately following provides evidence for a temporal reading. Here the sentence is effectively disambiguated to the temporal reading, so the unrestricted race model's penalty for disambiguation is expected to apply. In contrast, when the following syllable is unstressed, the race is able to run through to completion and one of the two parses is adopted. As medial verb stress is roughly compatible with either prosodic realization (lapse avoidance notwithstanding), the sentence remains locally ambiguous, so the parser is able to proceed without difficulty regardless of which grammatical commitment has been made. In fact, taking into account the possibility of a weak preference to avoid lapse offers a possible explanation for the unexpected interaction on first-pass reading times at the main verb. The verb was read significantly faster in C.2 than in T.2, a finding at odds with the established global temporal bias; however, the slowed reading times on T.2 may reflect an underlying race process which takes longer to complete due to the dispreferred rhythmic lapse that results, while a preceding contextual bias for the comparative reading yields a faster time to completion for the prosodically congruous alternating rhythm in condition C.2. Reading times for the initial stress conditions C.1 and T.1 are slower than C.2 and statistically indistinguishable from each other and condition T.2, a finding that is compatible with difficulty stemming from implicit prosodic disruption of the race process.

The second assumption in this account is that the early effect of verb stress upon regression probability

⁶ This argument for prosodic disambiguation draws heavily upon the reasoning of cue validity, positing that stress is a reliable cue to local structure. Similar reasoning has been used by advocates of the constraint-satisfaction approach to explain apparent asymmetries in the timing of different constraints upon parsing; for example, Spivey, Fitneva, Tabor, and Ajmani account for the delayed appearance of thematic role effects relative to those of lexical subcategorization by noting that "rather than becoming operative at an earlier point in time, subcategorization information may simply provide a probabilistically stronger constraint on grammaticality than thematic role information does" (2002, 219). Such a logic could be used to account for the earlier and stronger effects of implicit prosody within a constraint-based framework, assuming it is probabilistically more valid as a cue to structure than discourse context; however, this account fails to capture the direction of the effect, namely greater processing difficulty when prosody is compatible with the globally preferred reading. The penalty for disambiguation asserted by the unrestricted race model is thus more compatible with the current data.

at *mehr*, the ambiguous word, reflects parafoveal access to the initial syllable of the critical verb that immediately follows. This assumption is independently motivated in previous work by Ashby and Martin (2008), among others. Each main verb in the current experiment began with a 2-4 character prefix which is either obligatorily stressed or obligatorily unstressed according to the morphology of German verbs; thus, visual access to the initial syllable alone was sufficient to determine the presence or absence of stress clash in each trial. Ashby and Martin (2008) found converging evidence from a lexical decision task and an ERP experiment indicating that the prosodic structure of an initial syllable is reliably accessed during parafoveal preview. The parafoveal interpretation of the verb stress effect on *mehr* is also consistent with the findings of Breen and Clifton (2011), who found evidence for parafoveally triggered implicit metrical reanalysis occurring on the prosodically and syntactically ambiguous word itself, while syntactic reanalysis in the absence of prosodic ambiguity occurred in the disambiguating region as expected.⁷

These two assumptions motivate and reinforce each other: parafoveal access would explain why verb stress effects appear before the verb itself, and prosody as disambiguation would explain the unexpected direction of the effect: more regressions observed when there is a prosodic bias consistent with the global preference for the temporal reading. Of the two, parafoveal access is probably more plausible, which is fortunate, because this assumption is also necessary under other analyses. To abandon the prosody-as-disambiguation interpretation and exclude the verb stress effect upon regressions at *mehr* as statistical noise would still leave the main effect of verb stress at the modal verb immediately prior to syntactic disambiguation to be explained. In the absence of any reason to assume a global processing cost associated with initial stress on the main verb, this effect is explicable only in terms of parafoveal access of *als* conflicting with an earlier, prosodically-driven temporal reading of *mehr*.

Moving forward with the assumptions and analysis outlined above, upon reaching *als*, syntactic disambiguation occurs in all conditions toward the globally dispreferred comparative reading, but for initial verb stress conditions this is actually the second time that disambiguation occurs. Whereas *nicht mehr* had earlier appeared to resolve to the temporal reading due to prosody, it is now reanalyzed as comparative; hence, processing difficulty is observed at *als* and the word immediately following. Context may be expected to show an effect at this region in initial verb stress conditions as well, as the unrestricted race model, like the garden-path model, includes a role for context in guiding reanalysis; however, this effect may not be very large, owing to the earlier analysis in favor of the temporal reading. In contrast, for the medial verb stress conditions, *als* represents the first point of disambiguation.

The penalty for disambiguation is thus anticipated here, and context is expected to guide reanalysis as well, likely to a larger degree than in the initial stress conditions, as there has been no prior resolution to the alternative reading.

This account successfully captures the pattern of results observed at the point of disambiguation.⁸ The continuing influence of verb stress is reflected in marginal main effects on two early measures of processing: first fixation probability at *als*, and regression probability at *als+1*. The additional cost of prosodic reanalysis on top of syntactic reanalysis (Bader, 1998; Breen & Clifton, 2011) is evident in the second-pass measure of re-reading probability, which shows reliable main effects of verb stress on each region from *mehr* up to and including *als*. The anticipated effects of context are also apparent at the second word of the disambiguating region, where both first fixation probability and re-reading probability show main effects of context in the predicted direction. Post-hoc pairwise comparisons within both of these dependent measures reveal the expected asymmetry of this main effect — in both instances, the effect is driven by a significant difference in the medial stress conditions, so that the contextual facilitation for C.2 relative to T.2 is much greater than that of C.1 relative to T.1.

Another possible interpretation of the results presented here posits a privileged role for lexical information in online processing. This concept is reflected in the modified constraint-based model proposed by Boland and colleagues (Boland & Cutler, 1996), which claims that lexical constraints determine the initial generation of syntactic and semantic structures, and other constraints then select between the generated alternatives. The early effects of lexical-level verb stress and later effects of discourse context in our experiment do appear to mirror the early effects of lexical frequency and later effects of context found in Boland and Blodgett's study of ambiguous noun-verb homographs (2001); however, insofar as the stress clash avoidance found in our study is a supralexical rather than strictly lexical phenomenon, the precise relation of our results to this model's predictions is difficult to determine.

⁷ Note that, without further investigation, one cannot rule out an alternative explanation for the parafoveal effect: mislocated fixations (Drieghe, Rayner, & Pollatsek, 2008). Participants could have intended to fixate the main verb but saccadic error could result in their eyes falling on the preceding word; alternatively, measurement error in the tracker could be responsible for this effect. Given that the verb was skipped approximately 15% of the time (see below), these are possible explanations for the effect. We are grateful to a reviewer, Erik Reichle, for pointing this out.

⁸ Note, however, that the race model as originally proposed assumes that all sources of information are used immediately, so it predicts early effects of context. Thus, our assumption that context mainly has an effect on reanalysis is a departure from the original model.

Nonetheless, the core claim that lower-level cues which form bottom-up constraints upon syntactic structure may have a special role in parsing is compatible with our findings (cf. Snedeker & Yuan, 2008, for a related account of processing in development).

Regardless of whether the unrestricted race analysis presented above is ultimately borne out, the pervasive effects of implicit prosody found in both early and late stage dependent measures in the eye-movement record point to the need for new theoretical frameworks, ones which admit the possibility of implicit prosodic influence upon the earliest stages of syntactic parsing. The consistent main effects of verb stress in re-reading probability are particularly striking. While the effects of context observed following *als* suggest that global discourse bias does play a role in guiding reanalysis, implicit prosody emerges as the dominant factor in readers' decisions to re-visit the ambiguous region. These results speak against the possibility that a global contextual bias (at least, as construed in the present study) can eliminate a local prosodic garden path. When it comes to parsing syntactic structure, it appears that prosody not only follows syntax, but sometimes leads — even when this means leading in silence.

References

- Altmann, G., Nice, K. van, Garnham, A., & Henstra, J. (1998). Late closure in context. *Journal of Memory and Language*, 38(4), 459–484.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191–238.
- 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.
- Augurzky, P. (2006). *Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing*. MPI for Human Cognitive and Brain Sciences.
- Bader, M. (1996). Prosodic effects and the distinction between primary and non-primary phrases. In *Poster presented at the Conference on Architectures and Mechanisms of Language Processing (AMLaP)*. University of Torino, Torino.
- Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. *Reanalysis in sentence processing*, 1–46.
- Bates, D., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using Eigen and R syntax*.
- Binder, K., Duffy, S., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, 44(2), 297–324.
- Boland, J. (2004). Linking eye movements to sentence comprehension in reading and listening. In C. Clifton & M. Carreiras (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 51–76). New York, NY: Psychology Press.
- Boland, J., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45(3), 391–411.
- Boland, J., & Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the great divide. *Cognition*, 58(3), 309–320.
- Breen, M., & Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64(2), 153–170.
- Britt, M. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language; Journal of Memory and Language*.
- Carver, R. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, 84–95.
- 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 R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (chap. 15). Elsevier.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. *Natural Language Parsing*, 320–358.
- Dilley, L., & McAuley, J. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59(3), 294–311.
- Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology*, 61(8), 1239–1249.
- Fodor, J. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2), 285–319.
- Frazier, L., & Clifton, C. (1995). *Construal*. MIT Press.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, 32(2), 167–195.
- Hoeks, J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, 46(1), 99–119.
- Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147.
- Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of web experiments: A case study using the WebExp software package. *Behavior Research Methods*, 41(1), 1–12.
- Kelly, M., & Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 389.
- Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, 123, 1–20.
- Kjelgaard, M., & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2), 153–194.
- Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, 249–336.

- Logačev, P., & Vasishth, S. (2006). The em package for computing eyetracking measures. *University of Potsdam, Potsdam, Germany*.
- Logačev, P., & Vasishth, S. (2013). *A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension*. (Submitted)
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676.
- McDonald, J., Bock, K., Kelly, M., et al. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25, 188–188.
- McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Nespor, M., Vogel, I., et al. (1989). On clashes and lapses. *Phonology*, 6(1), 69–116.
- Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*, 11(3), 283–334.
- Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica*, 66(1-2), 95–112.
- R Development Core Team. (2006). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Available from <http://www.R-project.org> (ISBN 3-900051-07-0)
- R Development Core Team. (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1), 65–81.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 358–374.
- Rayner, K., Pacht, J., & Duffy, S. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language; Journal of Memory and Language*, 33, 527–544.
- Shih, S., Grafmiller, J., Futrell, R., & Bresnan, J. (2009). Rhythms role in genitive and dative construction choice in spoken English. In *31st Annual Meeting of the Linguistics Association of Germany (DGfS)*. University of Osnabrück, Germany.
- Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3), 238–299.
- Spivey, M., Fitneva, S., Tabor, W., & Ajmani, S. (2002). The time course of information integration in sentence processing. In P. Merlo & S. Stevenson (Eds.), *Sentence processing and the lexicon: Formal, computational, and experimental perspectives* (pp. 207–232). Amsterdam: J. Benjamins Publishing Company.
- Spivey, M., & Tanenhaus, M. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521–1543.
- Steinhauer, K., Alter, K., Friederici, A., et al. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191–196.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Traxler, M., Pickering, M., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39(4), 558–592.
- Van Gompel, R., & Pickering, M. (2007). Syntactic parsing. *The Oxford handbook of psycholinguistics*, 289–307.
- 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.
- Vasishth, S., von der Malsburg, T., & Engelmann, F. (2012). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*. Available from <http://dx.doi.org/10.1002/wcs.1209>
- Wickham, H. (2011). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.
- Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of the Conference on Spoken Language Processing*.