What eye movements can tell us about sentence comprehension

Shravan Vasishth,* Titus von der Malsburg and Felix Engelmann

Eye movement data have proven to be very useful for investigating human sentence processing. Eyetracking research has addressed a wide range of questions, such as recovery mechanisms following garden-pathing, the timing of processes driving comprehension, the role of anticipation and expectation in parsing, the role of semantic, pragmatic, and prosodic information, and so on. However, there are some limitations regarding the inferences that can be made on the basis of eye movements. One relates to the nontrivial interaction between parsing and the eye movement control system which complicates the interpretation of eye movement data. Detailed computational models that integrate parsing with eye movement control theories have the potential to unpack the complexity of eye movement data and can therefore aid in the interpretation of eye movements. Another limitation is the difficulty of capturing spatiotemporal patterns in eye movements using the traditional word-based eyetracking measures. Recent research has demonstrated the relevance of these patterns and has shown how they can be analyzed. In this review, we focus on reading, and present examples demonstrating how eye movement data reveal what events unfold when the parser runs into difficulty, and how the parsing system interacts with eye movement control. © 2012 John Wiley & Sons, Ltd.

INTRODUCTION

Consider the sentence: 'The key to the cabinets are on the table.' This sentence is ungrammatical (it should have been: 'The *key . . . is . . .*'), and yet the human sentence comprehension system (hereafter, the parser) often does not even register such an error.[1–3] Understanding the boundaries of the parsing mechanism—what errors it can and cannot detect, where it fails or experiences difficulty—has long been the subject of inquiry in sentence comprehension research.[4] Over the last 50 years, several issues have come under scrutiny, among them reanalysis processes,[5] the role of grammar in sentence comprehension,[6] incremental sentence processing and anticipation processes,[7,8] the nature of and constraints on dependency resolution (specifically, interference and retrieval cost, and expectation-based processing),[9–12] underspecification in parsing,[13–16] the role of silent and overt prosody in parsing,[17–19] and sentence comprehension disruptions in aphasics.[20,21]

Eyetracking has been found to be a rich source of information about parsing phenomena such as those listed above. The method, on the surface, seems relatively straightforward. Two widely used approaches are to record the eyes' fixation history (time spent on each word) as a sentence is read,[22] and to present subjects with a visual scene ('visual world')[7] and record their fixations on specific objects in the scene as they hear sentences. Both reading studies and the visual world paradigm have proven to be very informative tools for drawing inferences about sentence comprehension processes. Owing to space restrictions, we focus here on reading, and we discuss the kind of information eyetracking data can give us about processes triggered during sentence comprehension. As this article is intended for advanced researchers in sentence comprehension research, we assume some familiarity with the literature on human sentence processing. Detailed tutorials on sentence comprehension are available elsewhere.[23]

Eyetracking has the potential to inform us about *when* an event occurs in the parser (timing); *what* the parser does when it encounters difficulty (parsing events); and *how* attention, the eyes, and the parser interact (the eye–parser link).

*Correspondence to: vasishth@uni-potsdam.de

Department of Linguistics, University of Potsdam, Brandenburg, Germany

Regarding timing, whether a parsing event happens early or late after a word is recognized has important implications for theories of parsing. A good example is the grammar–parser relation: when faced with a reflexive pronoun like *himself/herself*, does the parser use the gender information available from the reflexive to initiate a search for the antecedent, does the parser use only syntactic constraints like Principle A of the binding theory,[24] or does it use both sources of information? The question this raises is: what role do grammatical constraints play in parsing, and what cues are used to complete dependencies? As discussed below, eyetracking can potentially answer questions of this kind.[6,25,26]

Related to the timing issue (the *when*), eyetracking also has the potential to uncover the *what*: the steps the parser takes when it encounters difficulty. In a classic paper, Frazier and Rayner investigated this question (discussed below in more detail).[5] Assuming a tight connection between parsing events and eye movements,[27] Frazier and Rayner argued that, when the parser is garden-pathed, the eyes are directed by the parser to the earlier region in the sentence where the initial misparse is then repaired. Subsequent research has suggested that although there is clearly some correspondence between parser actions and eye movements, the eyes probably do not move in perfect synchrony with the parser.[16,28–30] This loose coupling of the eye and parser complicates considerably the inferences that one can draw from eye movements but, in our opinion, the complexity of the eye–parser interaction is a hurdle only if we do not develop a detailed theory of the eye–parser connection.

Modeling the interaction between attention, the oculomotor system, and the parser (the *how*) is a major open question in sentence comprehension and has not yet received the attention it deserves. A large proportion of studies investigating parsing processes relies on eyetracking data,[4,22,31] but it is not until recently that the eye–parser connection has been taken up seriously.[32–35]

Before we can discuss these three points (the when, what, how), it is worth briefly reviewing the kind of dependent measures used in eyetracking, and what they might (or might not) reflect.

DEPENDENT MEASURES IN EYETRACKING

The dependent measures used in eyetracking are—in most cases, implicitly—intended to separate out early and late events. Thus, in reading studies it is quite normal to talk about 'early' measures versus 'late' measures; it is implied (or sometimes explicitly stated) that these early and late measures index early and late parsing events, respectively. Similarly, in visual world studies, how quickly the eyes move to a particular part of a scene is taken as an indicator of when a parsing event happened after hearing a string of words.[7] In reading, 'early' measures are those that involve first contact with a word or region as the reader approaches the word/region from the default reading direction (in English, from the left); examples are first fixation duration, gaze duration, and possibly first-pass regression probability. By contrast, 'late' measures are those involving revisits to a region (for example, re-reading time).[4]

Regarding the connection between early/late measures in reading and early/late parsing events, Clifton, Staub, and Rayner have the following to say (emphasis ours):[4]

> The terms "early" and "late" may be misleading, if they are taken to line up directly with first-stage vs second-stage processes that are assumed in some models of sentence comprehension (Rayner, Carlson, & Frazier, 1983; Frazier, 1987). Nonetheless, careful examination of when effects appear may be able to shed some light on the underlying processes. *Effects that appear only in the "late" measures are in fact unlikely to directly reflect first-stage processes; effects that appear in the "early" measures may reflect processes that occur in the initial stages of sentence processing*, at least if the measures have enough temporal resolving power to discriminate among distinct, fast-acting, processes.

We would like to add two comments to the above observation. First, if an effect shows up only in late measures, it may not always be reasonable to conclude that this reflects late events and not early events. One reason for this uncertainty is that early and late measures are generally highly correlated; this implies that late measures could in principle reflect early events regardless of whether or not the effect is statistically significant in early measures. One can imagine several reasons why an early parsing event could show statistical significance only in late measures. As mentioned above, we do not yet understand the eye–parser link as well as we should, and as a consequence the effects of the parser's events on eye movements are difficult to disentangle from oculomotor constraints. For example, the eye may move on from a critical region to a later word because of low-level oculomotor constraints, and the parser's actions may only later have an opportunity to influence eye movements, e.g., through re-reading of the critical region. Owing to such a possible lag, late measures might show an effect that starts early but becomes clearer only in late measures. One way to better understand the effect of

this lag is by developing more comprehensive models of reading where the interaction of oculomotor and parsing processes is well-specified. This appears to be a major open question in sentence processing research, but the first steps in this direction have already been taken.[32,34,35] Another way to determine whether an effect emerging only in late measures is truly due to a late parsing event is to increase power and determine whether the late-measure effect appears in early measures. We discuss an example below.[36]

A second comment regarding the above quote from Clifton and colleagues is about what we can conclude from early measures. One important issue in reading is spillover of processing from preceding regions. The duration of a first fixation on a particular critical word/region may reflect (inter alia) parsing costs that reflect processing of a pre-critical region. This may not be a concern when the precritical regions are identical across conditions; indeed, most eyetracking researchers take care to ensure that precritical regions are identical (or relevant control conditions are used). But how much material has to be identical to avoid confounding effects is unclear, and there may not be a general answer to that question. When the precritical regions cannot be kept identical in all conditions,[37] it can be difficult to draw reasonable inferences. In these cases, statistical tools can help to minimize the impact of potential spillover confounds.[9,16,38] The situation is complicated further by parafoveal preview, which may give the processing of a word a head start even before it is fixated.[39] Which aspects of a word (letter level, lexical, discourse information, etc.) can be processed before it is fixated is a matter of debate that makes it difficult to draw precise inferences about the time-course of word-processing. A more principled approach to deal with all these issues would be to develop a more complete specification of the eye–parser connection, and to investigate the relationship between the observed behavior (reading times, saccade targets) and the various visual and linguistic properties of the sentence through simulation. Is the increased reading time at a particular word due to lexical retrieval of the word or due to syntactic or other difficulty arising at a previously processed word?[40] Computational models of eye–parser interaction could help to answer questions like these.

A final, trivial but often neglected difficulty with the interpretation of eye-tracking measures along the temporal dimension stems from the inaccuracy of eye-tracking machinery. While a fixation may have been exactly on a critical word, spatial noise in the eye signal may give the impression that the fixation was on an earlier or later word. An effect can therefore be smeared across several words and may appear earlier

and/or later than it should. According to the technical specifications of many eyetracking systems, spatial noise should be reasonably low, but in practice the precision depends on many factors such as the font size used in the stimulus, the accuracy of the calibration procedure, visual aids, etc. Unfortunately, it is not common to report objective measures of the actual precision of the signal and most inferences simply assume that spatial noise does not play a notable role. We expect that reporting standards will change in the future; the first steps in this direction have already been taken (see http://www.cogain.org/info/eye-data-quality).

In sum, both early and late measures do have the potential to inform us about early and late parsing events, but, as Clifton and colleagues also point out, the mapping is far from obvious; this fact is occasionally forgotten in the literature. The interaction between oculomotor constraints and parsing events clearly needs a thorough investigation. Concurrent recordings of eye movements and electric brain potentials may become an exciting new source of evidence in this enterprise.[41–43]

We turn next to three examples that illustrate the informativeness of eyetracking data in developing sentence comprehension theories.

THE EYES REVEAL THE TIME-COURSE OF EFFECTS

A good example of the usefulness of reading studies for understanding timing issues is the work by Sturt.[6] He investigated whether principle A of the binding theory[24] is used at an early stage in determining the antecedent of a reflexive. In one of his eyetracking experiments, Sturt showed subjects texts like (1).

(1) Jennifer was pretty worried at the City Hospital. The surgeon who treated Jennifer/Jonathan had pricked himself/herself with a used syringe needle. There should be an investigation soon.

In the second sentence, the antecedent of *himself/herself* is always *surgeon*. Surgeons can certainly be women, but English has a stereotypical bias toward male surgeons. Sturt argues that if the parser initially uses *only* grammatical constraints (here, Principle A) to complete the dependency between the antecedent and reflexive, there should be no confusion (at least not initially) between the correct antecedent and any distractor noun phrase such as *Jonathan* in the case of the reflexive *himself*: this distractor noun phrase is inside a relative clause that modifies the correct antecedent, and if Principle A is used to filter out

irrelevant distractors, this noun phrase should not be considered when an antecedent-reflexive dependency is made. Thus, if only grammatical cues are used initially for antecedent search and the gender marking on the reflexive is ignored, the search should complete very early and without distraction from the noun phrase inside the relative clause.

Sturt found no effect in early dependent measures of gender match of the reflexive with the distractor noun, but in a late measure he found an effect. Sturt concluded that the parser carries out antecedent search using a syntactic constraint, initially ignoring the nonsyntactic cue of gender match between the distractor noun phrase and the reflexive, and only later on does gender match cause disruptions. This is the kind of situation that Clifton and colleagues refer to when they say (see above quote) that seeing an effect only in late measures may mean that the underlying event occurs late. In Sturt's experiment, this inference could indeed be correct.

However, the alternative possibility cannot be ruled out; Sturt's absence of effects in early measures are null results, and these could simply be due to low statistical power (equivalently, high Type II error probability). Repetitions of experiments that consistently fail to find an effect may not be convincing if power is low, i.e., if the probability of detecting an effect, given that the effect exists in nature, is low.[44] It is easy to test the claim that lack of power may be a problem for such inferences. For example, two eyetracking experiments[36] involving Mandarin Chinese reflexives were carried out where sample size was 130 and 150 participants, respectively (cf. Sturt's 24); these samples sizes were computed before running the experiment to achieve power of at least 0.80. In other words, conditional on the assumption that distractor effects occur and have a given magnitude (in terms of differences in reading time) and a given standard deviation for the differences (these can be estimated from a computational model or from existing data, for example), a power of at least 0.80 means that the chances of detecting such an effect are at least 80%. For example, if an expected difference between means is 30 ms, with standard deviation 75 ms, a power of 0.85 can be achieved in a two-sided paired *t*-test if we use 60 participants.

These experiments on Mandarin found an effect at the reflexive of a distractor noun in early measures. Of course, now one could argue that this is a Type I error! However, the key point here is that, in the above example, the probability of a Type I error (the probability of rejecting the null when it is in fact true) is 5%, whereas power (the probability of rejecting the null hypothesis when it is in fact false) is greater than 80%. For a particular experiment, that is as good as it gets: standardly used frequentist methods cannot give us more certainty than that. The best way to achieve more clarity on the question is to try to replicate the effect using high-power studies, and/or use more sophisticated methods of data analysis.[45,46]

Currently, we do not have enough data to resolve the research question discussed above (i.e., Sturt's conclusions may still be correct). But in this specific example, we argue that it is in principle possible to detect distractor effects in early measures. While the mechanisms subserving antecedent retrieval will continue to be a matter of debate, these results suggest that the absence of early effects in previous studies may not necessarily entail that the early stages of antecedent retrieval are immune to interference from distractors.

THE EYES CAN REVEAL PARSER STEPS

What does the parser do when it runs into trouble? Eyetracking has the potential to unpack parsing events, effectively leaving a historical record of what happened once the parser encounters difficulty. In an important paper,[5] Frazier and Rayner examined late closure and minimal attachment phenomena using reading. Their main question was: what can the *patterns* of eye movements tell us about parsing events? They used sentences such as the following:

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

The central assumption in this work was that the eyes would directly reflect reanalysis processes: if the parser builds an incorrect structure at the ambiguous region *jogs a mile* (i.e., if the parser attaches a *mile* as an object of *jogs*), upon detecting the error at *seems*, it could intelligently return to the ambiguous region to correct the misparse (i.e., a *mile* would become the subject of a new clause). Following Just and Carpenter,[27] Frazier and Rayner assume that a targeted revisit of the parser to the ambiguous region should be reflected in eye movement patterns. Specifically, the eyes would move directly from the disambiguating region to the ambiguous region reflecting the correction of the misparse. Frazier and Rayner named this selective reanalysis. The term selective reanalysis refers to the parsing event of targeted repair (as opposed to complete reparsing of the sentence,[47] or backtracking); it is an additional assumption that the targeted repair leaves a trace in the gaze trajectory. Frazier and Rayner visually inspected the scanpaths produced by subjects while reading garden-path sentences, and found that

subjects did indeed have a tendency to revisit the ambiguous region, suggesting that they were engaged in selective reanalysis (they also found evidence for re-reading, but these were considered to be instances of sentence-final wrap up effects as they were triggered from the sentence-final region). Later work (also a reading study) found supporting evidence for selective reanalysis in Spanish attachment ambiguities.[28]

It is reasonable to question the assumption that the eye and parser are tightly linked. Mitchell and colleagues[29] raised the objection that the eye movement pattern in the Frazier and Rayner study that was consistent with selective reanalysis could simply be a consequence of the eyes' tendency to move leftward by one word as soon as processing difficulty is encountered—perhaps in order to buy time for the parser before proceeding to new material. In Frazier and Rayner's sentences (see example 2 above), the ambiguous region happened to be adjacent to the disambiguating region, making it impossible to conclude what the driver of the regressions to the ambiguous region was. Was it selective reanalysis, or just a short regression to avoid taking in more information to the right while re-parsing is done 'in place', at the disambiguating region (Mitchell and colleagues call this 'time outs')? Mitchell and colleagues ran further studies to separate these two possibilities out (among other things, they increased the distance between the ambiguous and disambiguating regions), and found evidence for both parser-directed regressions to the ambiguous region (i.e., evidence consistent with the original selective reanalysis findings), but also evidence for their Time Out hypothesis.

This discussion about the eye–parser relationship in the face of a misparse requires examining scanpaths rather than traditional fixation duration measures. The three papers mentioned above rely either on visual inspection of scanpaths,[5] or on regressive landing site distributions (the distribution of the landing sites of the first regressive saccade after a disambiguating region is reached).[28,29] Due to the complexity inherent in the study of scanpaths, the analysis of landing sites of a single regressive saccade represents a simplified way for tackling the question: which patterns does the eye movement record show? Scanpaths are difficult to characterize because not only do they consist of a sequence of fixations but also each fixation has a particular fixation duration too; thus, defining a metric for scanpath similarity requires taking not only position but also duration into account. However, recent work has shown that it is possible to study scanpaths by quantifying their relative similarity to each other and then using clustering techniques to identify characteristic scanpaths.[16,30]

Von der Malsburg and Vasishth developed a suitable similarity measure for scanpaths and reexamined Meseguer and colleagues' Spanish data where evidence had been found for selective reanalysis. In this analysis, they found an interesting additional detail when scanpaths were clustered to reveal characteristic patterns.[30] Two characteristic patterns were re-reading of the sentence after the disambiguating point was reached, and short leftward saccades of the type that Mitchell and colleagues call Time Out regressions (see Figure 1). These patterns were later replicated in a follow-up study on Spanish.[16] Interestingly, these patterns could not be identified using the traditional word-based reading measures discussed above. Thus, the scanpath analysis showed that re-reading in the face of a misparse is a common strategy.[47] In fact, as mentioned earlier, the original study by Frazier and Rayner had found evidence for re-reading, but these were considered to be instances of sentence-final wrap-up effects (because in that study the re-reading pattern had its onset at the final word). The follow-up study mentioned above[16] found re-reading being triggered even from non-sentence final positions, suggesting that re-reading may reflect a fresh re-start by the parser once a parsing error is detected. As Lewis points out,[47] re-reading is not an optimal strategy in terms of time, but it costs little by way of
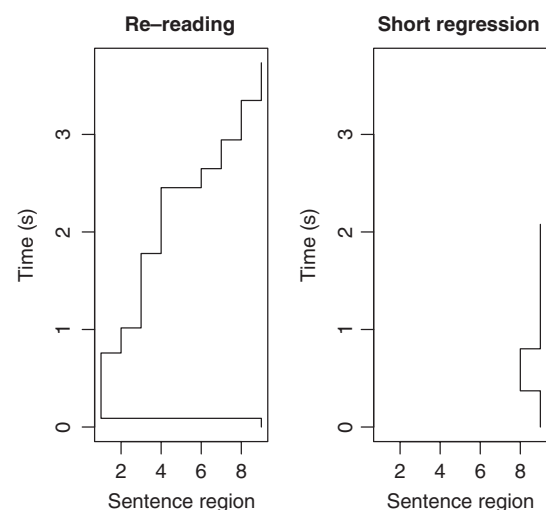
FIGURE 1 | This figure shows two characteristic scanpath patterns found by von der Malsburg and Vasishth in the Meseguer, Carreiras, and Clifton Spanish reanalysis dataset. One scanpath pattern is a complete re-start (re-reading) after the disambiguating region[9] is encountered. The other scanpath pattern was a short regression to the immediately preceding word; this is consistent with the suggestion by Mitchell and colleagues of a time-out—as the parser reanalyses, the eyes are prevented from moving forward, and as a result the eyes either stay on the current word or just revisit the preceding word until syntactic processing is finished. See text for details.

memory since earlier choice points for the parser do not have to be memorized and recalled.

The Spanish replication study[16] also measured working memory capacity of subjects. It revealed that subjects with higher working memory capacity tended to make attachment decisions more often than low-capacity subjects; this had the consequence that high-capacity subjects made more attachment errors than low-capacity subjects, which was reflected in higher rates of regressive eye movements. This finding is consistent with the good-enough account of sentence comprehension which posits that the parser, in order to save resources, does not always aim for complete comprehension and faithful representations.[13,14] These results serve as a demonstration of the usefulness of studying scanpaths because the key effects of working memory were revealed only when the trials were separated according to the eye movement pattern that occurred.

Scanpath analyses can also recover variation in eye movement patterns due to age-related differences among readers and due to syntactic complexity of the sentence material.[48,49] This was shown in an analysis of the Potsdam Sentence Corpus[50] which contains eye movements for young and old readers. The principal findings were that greater syntactic parsing difficulty results in more irregular scanpaths; older readers showed more irregular scanpaths than younger readers; and older readers showed a smaller effect of syntactic parsing difficulty (perhaps disruptions due to impaired executive control mask effects of syntax in older readers).[51]

In sum, eyetracking can be very informative about the parsing steps taken after an error or difficulty occurs, but the inferences that can be drawn depend on the assumptions one makes about eye movement patterns directly reflecting underlying cognitive processes. These assumptions are often left implicit in the literature.

HOW DO THE EYES AND PARSER INTERFACE?

As we have mentioned throughout this review, the specific interaction between the eyes and the parser is less than clear. Eye movements are subject to oculomotor constraints that have nothing to do with parsing. These oculomotor constraints play a crucial role in deciding how long the eyes fixate on a word, and where they move to next. This fact complicates in particular the interpretation of early measures and saccade targets.

One response to this issue has been to avoid doing reading studies altogether; many researchers use visual world data exclusively to study sentences comprehension. This might also be because the variance explained by factors considered to belong to higher level cognition can be quite low, e.g., in eyetracking corpus studies.[52] However, variance explained has little relevance in controlled experimental designs; it does not help in determining the theoretical importance of an effect.[53,54]

In any case, one point that everyone would generally agree on is that the interaction of oculomotor processes and parsing processes should itself be an object of study, not only just for reading studies but also for the visual world paradigm. A functionally complete model that defines the eye–parser connection fully could serve as a vital tool for disentangling the influences of oculomotor constraints and cognition.

The eye movement control community has already made significant advances in developing detailed models of how oculomotor processes affect reading.[55–60] Although assumptions of the major models differ—e.g., E-Z Reader is a serial attention model, whereas SWIFT allows parallel processing of words guided by an attentional gradient—these models provide a very good specification of the 'low level' (from a linguistic point of view) processes driving the eyes during reading. Besides a specification of motor planning and execution these models already incorporate linguistic word-level information like word frequency, length, and predictability which enables them to predict early fixation measures with high accuracy.

The first serious attempt to extend this class of models to cover parsing-related processes involved E-Z Reader.[32] In its latest version, this model has a new module that stands in for postlexical processes (i.e., integration of the word in its context). The particular commitment about the eye–parser connection the model makes is the following: the only impact of the parser on eye movements is the possibility to interrupt 'normal' reading guided by low-level factors; this interruption triggers regressions and refixations when processing difficulty increases. What is missing is a theory of postlexical parsing processes (E-Z Reader specified a proxy for such a theory), but in principle this architecture could be extended to study the eye–parser connection.

In related work, Engelmann and colleagues have developed a framework within the ACT-R architecture for cognitive modeling.[34,35,61] ACT-R already incorporates an eye movement control model, EMMA,[62] which is a simplified version of E-Z Reader, generalized to be applicable in various cognitive tasks. The authors connected this eye movement model with
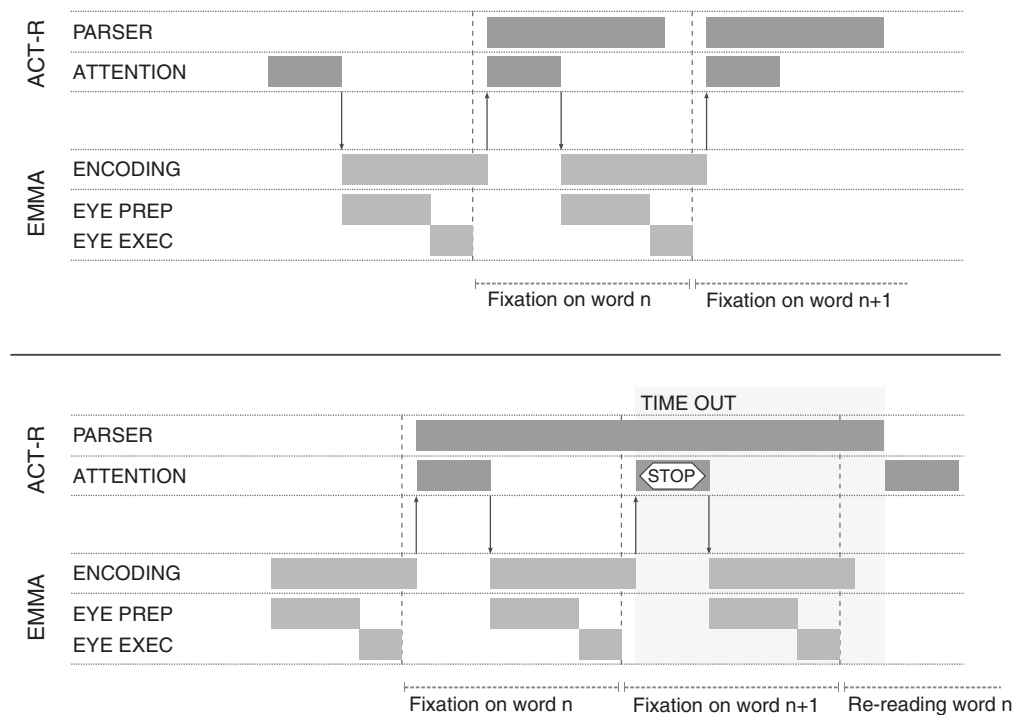
FIGURE 2 | This schematic figure shows how the ACT-R-based parsing architecture of Lewis and Vasishth interfaces with the EMMA eye movement control model. The upper panel shows the uninterrupted reading process. The only ACT-R rule that interacts with eye movement control in a top-down way is the shift of attention. As soon as an attention shift is requested (ATTENTION) the eye movement module starts the word recognition process (ENCODING) and at the same time programs a saccade to the same word. The preparation stage of the saccade programming (EYE PREP) can be canceled by an upcoming attention shift, which leads to a skipping of the targeted word. Once the beginning of the execution stage (EYE EXEC) has passed, an eye movement will be carried out inevitably. The completion of the attention shift, which includes the recognition of the word, is the signal for the parsing module (PARSER) to begin the integration into the syntactic structure. This includes the creation of new syntactic nodes, the retrieval of previously created structural chunks from memory, and finally the grammatical combination of both. While the parser is carrying out these steps attention is shifted to the next word and a new saccade is programmed. The time needed to retrieve an item from memory varies as a function of decay over time and similarity-based interference. Consequently, dependent on the syntactic configuration of the sentence it is possible that the structural integration of a word is still in process while the recognition of the next word has already completed. This scenario is shown in the lower panel. In this situation, the next word naturally cannot be integrated yet. Instead a Time Out rule fires, which initiates an attention shift to the left of the current word in order to buy time for the integration process to finish.

an implemented theory of parsing[25,38,63,64] which uses the ACT-R mechanisms of memory retrieval. Following the advances made in E-Z Reader, the model was extended to initiate a regression when the parsing module is not ready to integrate the next word. See Figure 2 for an illustration of this model. This simple interaction principle makes it possible to investigate later measures like regression probability and re-reading time, which arise as the result of a well-defined connection between a parsing theory and eye movement control.

Another modeling approach that investigates how eye movements arise from the interaction of oculo-motor constraints and language processing assumes that language knowledge is used to aid word-identification under noisy visual input.[33]

In this model, eye movement strategies arise from a process optimizing a tradeoff of reading speed and word recognition accuracy. In a nutshell: Word identities are guessed based on the input and the language model. If a word is guessable using only parafoveal visual input, it is skipped. Regressions are triggered when the confidence in earlier guesses falls. Although, this model does not explicitly model parsing processes, it can be used to investigate how language processing and oculo-motor constraints together determine eye movements in reading.

Eye movement models like those mentioned above have the potential to change the way we evaluate the predictions of theories of sentence comprehension. The inferences we can draw from regression-related measures like regression probability and re-reading time particularly depend on our assumptions about the coupling of the eye and parser. Studies based on eye movement corpa show that these measures are highly related to quantitative

metrics of parsing difficulty.[65] But only with detailed computational models is it possible to uncover the specific underlying drivers of regressive eye movements. In their current state, E-Z Reader as well as the ACT-R model of Engelmann et al. produce short regressions just for the purpose of a 'time out'; there is no linguistically guided regression target.[29] As the ACT-R model is equipped with a fully specified parser, it can easily be used to model long-range regressions and their relation to the structural repair processes that are necessary in a garden-path situation. With models of this kind it has become possible to study in detail the link between scanpath patterns and parser actions.

CONCLUSION

Eyetracking data can be very informative for evaluating theories of sentence comprehension, although care is needed regarding the interpretation of dependent measures. To some extent, the information available in eyetracking data has not yet been fully exploited, as is seen from the recent work on scanpaths in reading; we expect that future research will develop further novel-dependent measures in order to extract more empirical detail about the eye–parser connection. A related point is that the interaction of oculomotor constraints and parsing constraints needs to be understood better. We expect that the further development of models of eye–parser interaction will be a crucial future direction for sentence comprehension researchers. We believe that this line of research—the use of modeling to study the eye-parser connection, and the use of advanced methods such as scanpath analyses—will provide important new insights for sentence comprehension theories.

REFERENCES

1. Bock K, Miller C. Broken agreement. *Cognit Psychol* 1991, 23:45–93.

2. Pearlmutter NJ, Garnsey SM, Bock K. Agreement processes in sentence comprehension. *J Mem Lang* 1999, 41:427–456.

3. Wagers M, Lau E, Phillips C. Agreement attraction in comprehension: representations and processes. *J Mem Lang* 2009, 61:206–237.

4. Clifton C, Staub A, Rayner K. Eye movements in reading words and sentences. In: Gompel RV, Fisher M, Murray W, Hill RL, eds. *Eye Movements: A Window on Mind and Brain.* Amsterdam: Elsevier; 2007.

5. Frazier L, Rayner K. Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cogn Psychol* 1982, 14:178–210.

6. Sturt P. The time-course of the application of binding constraints in reference resolution. *J Mem Lang* 2003, 48:542–562.

7. Kamide Y, Altmann G, Haywood S. The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J Mem Lang* 2003, 49:133–156.

8. Levy R. Expectation-based syntactic comprehension. *Cognition* 2008, 106:1126–1177.

9. Bartek B, Lewis RL, Vasishth S, Smith M. In search of on-line locality effects in sentence comprehension. *J Exp Psychol Learn Mem Cogn* 2011, 37:1178–1198.

10. Van Dyke J. Interference effects from grammatically unavailable constituents during sentence processing. *J Exp Psychol Learn Mem Cogn* 2007, 33:407.

11. Vasishth S, Drenhaus H. Locality effects in German. *Dialogue Discourse* 2011, 1:59–82.

12. Levy R. Keller F. Expectation and locality effects in German verb-final structures. *J Mem Lang*. In press.

13. Sanford AJ, Sturt P. Depth of processing in language comprehension: not noticing the evidence. *Trends Cogn Sci* 2002, 6:382–386.

14. Ferreira F, Patson N. The good enough approach to language comprehension. *Lang Linguist Compass* 2007, 1:71–83.

15. Swets B, Desmet T, Clifton C, Ferreira F. Underspecification of syntactic ambiguities: evidence from self-paced reading. *Mem Cognit* 2008, 36:201–216.

16. von der Malsburg T, Vasishth S. Scanpath patterns in reading reveal syntactic underspecification and reanalysis strategies. *Lang Cogn Process*. In press.

17. Ito K, Speer S. Anticipatory effects of intonation: eye movements during instructed visual search. *J Mem Lang* 2008, 58:541–573.

18. Breen M, Clifton C. Stress matters: effects of anticipated lexical stress on silent reading. *J Mem Lang* 2011, 64:153–170.

19. Kentner G. Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition* 2012, 123:1–20.

20. Dickey M, Choy J, Thompson C. Real-time comprehension of wh-movement in aphasia: evidence from eye-tracking while listening. *Brain Lang* 2007, 100:1–22.

21. Hanne S, Sekerina I, Vasishth S, Burchert F, Bleser RD. Chance in agrammatic sentence comprehension: what does it really mean? Evidence from eye movements of German agrammatic aphasics. *Aphasiology* 2011, 25:221–244.

22. Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 1998, 124:372–422.

23. Frazier L. Sentence processing: a tutorial review. In: Coltheart M, ed. *Attention and Performance XII: The psychology of reading*. Hillsdale, NJ: Erlbaum; 1987, 559–586.

24. Chomsky N. Lectures on Government and Binding, Foris, Dordrecht, 1981.

25. Patil U, Vasishth S, Lewis RL. Retrieval interference in syntactic processing: the case of reflexive binding in English. Manuscript under revision.

26. Dillon B, Structured Access in Sentence Comprehension. PhD Thesis, University of Maryland, MD, USA, 2011.

27. Just M, Carpenter P. A theory of reading: from eye fixations to comprehension. *Psychol Rev* 1980, 87:329–354.

28. Meseguer E, Carreiras M, Clifton C, Jr. Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Mem Cognit* 2002, 30:551–561.

29. Mitchell D, Shen X, Green M, Hodgson T. Accounting for regressive eye-movements in models of sentence processing: a reappraisal of the selective reanalysis hypothesis. *J Mem Lang* 2008, 59:266–293.

30. von der Malsburg T, Vasishth S. What is the scanpath signature of syntactic reanalysis? *J Mem Lang* 2011, 65:109–127.

31. Rayner K. Eye movements in reading: models and data. *J Eye Mov Res* 2009, 2:1–10.

32. Reichle E, Warren T, McConnell K. Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bull Rev* 2009, 16:1–21.

33. Bicknell K, Levy R. A rational model of eye movement control in reading. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Uppsala, Sweden, 2010.

34. Engelmann F, Vasishth S, Engbert R, Kliegl R. A framework for modeling the interaction of syntactic processing and eye movement control. Submitted for publication.

35. Engelmann F, Vasishth S, Engbert R, Kliegl R, A framework for modeling the interaction of syntactic processing and eye movement control. In Proceedings of the Annual CUNY Sentence Processing Conference. New York, NY, 2012.

36. Jager L, Vasishth S, Interference effects in anaphor resolution: evidence from Mandarin. In: *Proceedings of the AMLaP Conference*, Trento, Italy, 2012.

37. Birch SL, Rayner K. Linguistic focus affects eye movements during reading. *Mem Cognit* 1997, 25:653–660.

38. Vasishth S, Lewis RL. Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 2006, 82:767–794.

39. Rayner K. The perceptual span and peripheral cues in reading. *Cogn Psychol* 1975, 7:65–81.

40. Boston MF, Hale JT, Vasishth S, Kliegl R. Parallel processing and sentence comprehension difficulty. *Lang Cogn Process* 2011, 26:301–349.

41. Kretzschmar F, Bornkessel-Schlesewsky I, Schlesewsky M. Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport* 2009, 20:1613–1618.

42. Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R. Coregistration of eye movements and EEG in natural reading: analyses and review. *J Exp Psych General* 2011, 140:552–572.

43. Kliegl R, Dambacher M, Dimigen O, Jacobs A, Sommer W. Eye movements and brain electric potentials during reading. *Psychol Res* 2012, 76:145–158.

44. Xiang M, Dillon B, Phillips C. Illusory licensing effects across dependency types: ERP evidence. *Brain Lang* 2009, 108:40–55.

45. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. 1st ed. Boca Raton, FL: Chapman and Hall/CRC; 1995.

46. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press; 2007.

47. Lewis RL. Reanalysis and limited repair parsing: leaping off the garden path. *J Psycholinguist Res* 1998, 247–284.

48. von der Malsburg T, Kliegl R, Vasishth S. Determinants of scanpath regularity in reading. Submitted for publication.

49. von der Malsburg T, Kliegl R, Vasishth S. Determinants of scanpath regularity in reading. In: *Proceedings of the CUNY Conference on Sentence Processing*, New York City, NY, USA, 2012.

50. Kliegl R, Nuthmann A, Engbert R. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J Exp Psychol General* 2006, 135:12–35.

51. Wotschack C, Kliegl R. Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Q J Exp Psychol*. In press.

52. Kliegl R, Olson R, Davidson B. Regression analyses as a tool for studying reading processes: Comment on Just and Carpenter's eye fixation theory. *Mem Cognit* 1982, 10:287–296.

53. Abelson R. A variance explanation paradox: when a little is a lot. *Psychol Bull* 1985, 97:129–133.

54. Prentice D, Miller D. When small effects are impressive. *Psychol Bull* 1992, 112:160–164.

55. Reichle E, Pollatsek A, Fisher D, Rayner K. Toward a model of eye movement control in reading. *Psychol Rev* 1998, 105:125–157.

56. Reichle E, Rayner K, Pollatsek A. The E-Z Reader model of eye-movement control in reading: comparisons to other models. *Behav Brain Sci* 2003, 26:445–476.

57. Engbert R, Nuthmann A, Richter E, Kliegl R. SWIFT: a dynamical model of saccade generation during reading. *Psychol Rev* 2005, 112:777–813.

58. Richter EM, Engbert R, Kliegl R. Current advances in SWIFT. *Cogn Syst Res* 2006, 7:23–33.

59. Legge GE, Klitz TS, Tjan BS. Mr. Chips: an ideal-observer model of reading. *Psychol Rev* 1997, 104:524–553.

60. Legge GE, Hooven TA, Klitz TS, Mansfield JS, Tjan BS. Mr Chips 2002: new insights from an ideal-observer model of reading. *Vision Res* 2002, 42:2219–2234.

61. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. *Psychol Rev* 2004, 111:1036–1060.

62. Salvucci D. An integrated model of eye movements and visual encoding. *Cogn Syst Res* 2001, 1:201–220.

63. Lewis RL, Vasishth S. An activation-based model of sentence processing as skilled memory retrieval. *Cogn Sci* 2005, 29:1–45.

64. Vasishth S, Brüssow S, Lewis RL, Drenhaus H. Processing polarity: how the ungrammatical intrudes on the grammatical. *Cogn Sci* 2008, 32:685–712.

65. Boston MF, Hale JT, Patil U, Kliegl R, Vasishth S. Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam Sentence Corpus. *J Eye Mov Res* 2008, 2:1–12.