Journal of Memory and Language 59 (2008) 266–293

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

# Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis

Don C. Mitchell *, Xingjia Shen, Matthew J. Green, Timothy L. Hodgson

*School of Psychology, University of Exeter, Exeter, Devon EX4 4QG, UK*

ARTICLE INFO

*Article history:*
Received 15 December 2005
Revision received 4 June 2008
Available online 30 July 2008

*Keywords:*
Syntactic ambiguity resolution
Human parsing
Regressions
Eye-movements
Selective Reanalysis
Computational models

ABSTRACT

When people read temporarily ambiguous sentences, there is often an increased prevalence of regressive eye-movements launched from the word that resolves the ambiguity. Traditionally, such regressions have been interpreted at least in part as reflecting readers' efforts to re-read and reconfigure earlier material, as exemplified by the Selective Reanalysis hypothesis [Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210]. Within such frameworks it is assumed that the selection of saccadic landing-sites is *linguistically supervised*. As an alternative to this proposal, we consider the possibility (dubbed the Time Out hypothesis) that regression control is partly decoupled from linguistic operations and that landing-sites are instead selected on the basis of low-level spatial properties such as their proximity to the point from which the regressive saccade was launched. Two eye-tracking experiments were conducted to compare the explanatory potential of these two accounts. Experiment 1 manipulated the formatting of linguistically identical sentences and showed, contrary to purely linguistic supervision, that the landing site of the first regression from a critical word was reliably influenced by the physical layout of the text. Experiment 2 used a fixed physical format but manipulated the position in the display at which reanalysis-relevant material was located. Here the results showed a highly reliable linguistic influence on the overall distribution of regression landing sites (though with few effects being apparent on the very first regression). These results are interpreted as reflecting mutually exclusive forms of regression control with fixation sequences being influenced both by spatially constrained, partially decoupled supervision systems as well as by some kind of linguistic guidance. The findings are discussed in relation to existing computational models of eye-movements in reading.



Over the last quarter of a century the field of psycholinguistics has arguably come of age with the development of a wide range of effective, fully-implemented computational models of processes ranging from lexical and oculomotor models of eye movement control in reading (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Pollatsek, Reichle, & Rayner, 2006; Reichle, Pollatsek, Fisher, & Rayner, 1998), through word reading and pronunciation (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; McClelland & Rumelhart, 1981; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), auditory word recognition (e.g. McClelland & Elman, 1986; Norris, 1994), speech production (e.g. Dell, 1986), thematic assignment (e.g., McClelland & Kawamoto, 1986; St John & McClelland, 1990) and aspects of language acquisition (Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986).

In contrast with several other areas, in the field of parsing the development of fully-quantified models has been comparatively tentative and it is only in recent years that we have seen the emergence of simulations that are capable of making detailed numerical predictions about the dominant on-line empirical phenomena in the field such

* Corresponding author. Fax: +44 1392 264623.
*E-mail address:* d.c.mitchell@exeter.ac.uk (D.C. Mitchell).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.06.002

as the word-by-word or phrase-by-phrase reading times at different points of a sentence (e.g., Christiansen & Chater, 1999; Christiansen & Chater, 2001; Green & Mitchell, 2006; Konieczny & Döring, 2003; Levy, 2008; Lewis, 1993; Lewis & Vasishth, 2005; MacDonald & Christiansen, 2002; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Narayanan & Jurafsky, 2002; Rohde, 2002; Spivey & Tanenhaus, 1998; Stevenson, 1993; Stevenson, 1998; Tabor, Juliano, & Tanenhaus, 1997).

In work on syntactic processing, the experimental evidence comes overwhelmingly from one or other of two on-line methods: self-paced reading (e.g., Altmann & Steedman, 1988; McRae et al., 1998) and eye-tracking (e.g., Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Of these two approaches, the measurement of eye-movements has been the focus of over a hundred published papers in the field (as detailed in an extensive recent review by Clifton, Staub, & Rayner, 2007). Given this wealth of empirical evidence, there are strong pragmatic grounds for using eye-tracking records as one of the primary forms of data for benchmarking and evaluating computational models. In practice, however, it turns out that it has been relatively rare for researchers to fit eye-tracking data. In total, we have only been able to locate seven published papers that use eye-tracking data as a basis for fitting the numerical predictions of computational models of syntactic processing (Binder, Duffy, & Rayner, 2001; Ferretti & McRae, 1999; Just & Carpenter, 1992; Konieczny & Döring, 2003; Spivey & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, & Hanna, 2000, and Vasishth, Brüssow, Lewis & Drenhaus, 2008). By way of comparison, there have been at least eleven attempts to model self-paced reading data (e.g., Elman, Hare, & McRae, 2005; Gibson, 1998; Grodner & Gibson, 2005; Hale, 2003; Just & Carpenter, 1992; Lewis & Vasishth, 2005; MacDonald & Christiansen, 2002; McRae et al., 1998; Narayanan & Jurafsky, 2002; Spivey & Tanenhaus, 1998; Tabor et al., 1997).

Given the small body of prior work on modelling eye-movement data (as driven by parsing), it is perhaps not surprising that there are many ground-clearing issues still to be resolved. In relation to parser-generated phenomena, there is as yet no consensus about which of the many standard eye-tracking measures should be made the subject of quantitative predictions. Nor is there any agreement about the nature of the interplay between the forms of control exercised at the syntactic level (and perhaps by other higher level linguistic operations) and those imposed by lexical effects and other similar "low-level" operations. Indeed, the coverage of prior work has been so sparse that there are numerous extensively studied parsing-linked eye-tracking effects that have never been captured by quantitative models. Many of these are linked to the phenomena associated with regressive eye-movements. It is these particular patterns that are the main focus of the present paper.

## The challenge of accounting for regressive eye-movements

It has long been known that people very rarely read sentences word by word in the "correct" order. Buswell (1922) reported that for fluent readers approximately 10% of eye-movements are characterized as being *regressive* in the sense that they move back to earlier material rather than passing on to the next unread part of the sentence. Modern estimates, if anything, show slightly greater prevalence of regressive eye-movements with regression rates often reported as being in the range of 10–15% (e.g., Rayner & Pollatsek, 1989, Chapter 4; Rayner, 1998). Based on an analysis of a large corpus of data collected from adults reading a novel, Vitu and McConkie (2000) reported that 15.3% of all saccades were regressive.

There is little doubt that regressions can be triggered by problems encountered at any of the various levels of linguistic analysis from graphemic processing and lexical analysis at one extreme to discourse processing at the higher level. However, it is of particular concern for present purposes that there is solid evidence that regressive eye-movements are associated with difficulties in syntactic processing. For example, regression rates have been shown to increase in the disambiguation regions of sentences (e.g., Frazier & Rayner, 1982; Meseguer, Carreiras, & Clifton, 2002; Rayner, Carlson, & Frazier, 1983; Traxler, Pickering, & Clifton, 1998; Trueswell, Tanenhaus, & Kello, 1993; van Gompel, Pickering, & Traxler, 2001). Partly as a consequence of this, regression-based measures of reading time also increase in such regions (e.g., Brysbaert & Mitchell, 1996; Desmet, De Baecke, & Brysbaert, 2002; van Gompel et al., 2001). Given the pervasiveness of these phenomena, it would be reasonable to expect detailed explanations of regressive eye-movements to feature prominently in the development and evaluation of implemented models of sentence processing. In practice, while one or two studies have taken on the task of modelling measures that *incorporate* regressive fixations (Binder et al., 2001; Konieczny & Döring, 2003), there seems to have been only one study that has responded to the challenge of predicting the prevalence of regressions at different points in a sentence (namely, Tanenhaus et al., 2000). We are not aware of any work at all that has set out to provide syntactically grounded computation-based predictions of the spatial distribution of regressive saccades during sentence processing. With the exceptions listed above and a very recent paper by Reichle, Warren, and McConnell (submitted for publication), all existing computational models of eye-tracking in reading explicitly restrict their machinery to operations below the level of syntactic processing (e.g., Engbert, Longtin, & Kliegl, 2002; Reichle, Rayner, & Pollatsek, 2003; Reilly & Radach, 2003). In effect this makes it impossible for such models to offer any insights into the nature of any interconnection there may be between parsing and eye-movements.

These gaps in our understanding of the full transmission system compromise our ability to use regression data to throw light on the nature of the syntactic operations themselves. While there is a substantial body of work examining the kinds of linguistic operation (termed "reanalysis") assumed to occur at points where regressions are prevalent (e.g., Ferreira & Henderson, 1991; Ferreira & Henderson, 1993; Fodor & Ferreira, 1998; Fodor & Inoue, 1994; Sturt & Crocker, 1998; Sturt, Pickering, & Crocker, 1999; van Dyke & Lewis, 2003), there has been remarkably little discussion of the ways in which eye-movements themselves might be orchestrated by such revision and repair operations (for notable exceptions to this general lack of scrutiny see Meseguer et al., 2002 and Frazier & Rayner, 1982). As far as we can establish, no one has even offered a theoretical explanation of the rather counterintuitive finding that first-pass reading times are shorter when terminated by regressive exits than when they end with a progressive movement to new material (first noted by Altmann, Garnham, & Dennis, 1992, and later corroborated by Rayner & Sereno, 1994). Indeed, in discussion of the linkages between eye-movements and linguistic operations, it has even been suggested (for example, by Lewis, 1998, p. 253) that there may be circumstances in which eye-movements are entirely decoupled from reanalysis operations. Where this occurs, attempts to reconfigure the input may manifest themselves not as *overt* patterns of eye-movements but more *covertly* as increased reading times at specific points in the sentence. Clearly if eye-movement are entirely disengaged from reanalysis in this way there would be little to be gained in any simulation by incorporating reanalysis as one of the drivers of oculomotor activity.

## Potential characteristics of eye-control systems

The present paper sets out to examine a range of different issues in relation to the syntactic control of eye-movements, and most particularly of regressive eye-movements. Two broad themes will emerge in the analysis that follows. The first revolves around the extent to which regressions are *linguistically supervised* rather than being triggered and then guided by non-linguistic operations. Much of our attention here will focus on the forms of control that might be imposed by systems responsible for the syntactic analysis of sentences. The second thread of our argument concerns the extent to which linguistic and other forms of eye-movement control are *tightly-coupled*, cutting out most forms of unpredictability in the system. Models of eye-movements often start with representations of the input (in the form of units standing for letters, words and spaces) together with some way of characterizing certain mental operations performed on the input (e.g., lexical processing). They then apply mathematical calculations to convert this raw material into an output consisting of quantities like the distance covered by the next saccade, the duration of current and future fixations and so on. At one extreme it is possible to imagine systems that are entirely deterministic, with all eye-movements being totally determined by the input. At the other extreme there may be no relationship at all between the inputs and movements. More realistically, one can conceive of systems that are *loosely-coupled* in the sense that they display a degree of randomness that co-exists with genuine constraints on the output. A little later we consider where the human linguistic eye-guidance system falls on this dimension from loose to tight control. In addition to these strands of discussion, we will also consider how control might be *shared* between different drivers, and how different views on this division of responsibility impact on methodological and theoretical issues. There is abundant evidence that eye-movements are influenced by a multiplicity of factors ranging from the spatial details of the print, through the characteristics of individual words on view, overall coherence of the text to the task demands being followed by the reader (from skimming to proofreading) (Rayner, 1998). What is less clear is how these different influences are coordinated. No individual driver can have a deterministic effect if there are circumstances in which it cedes control to other systems. It is vital to consider such interplay in any attempt to edge toward a realistic description of the overall control system.

More concretely, across a variety of different conceptions of linguistic supervision we will be considering the relative merits of a traditional, tightly-coupled view of the linguistic control of eye-movements and those of a more loosely-coupled alternative to this account. The traditional conjecture—not always explicitly spelt out—is that during reading both the placement of the eyes and the duration of fixations are largely determined by linguistic operations. An early expression of this view was termed the "eye-mind assumption" (Just & Carpenter, 1980). This states that "there is no appreciable lag between what is fixated and what is processed". Here, "what is processed" is not necessarily restricted to *linguistic* operations, but there is nevertheless a presupposition that there is a tight linkage between 'what is fixated' and some mental operation of interest. In a great deal of eye-tracking work, this linking assumption is tacitly used to sanction inferences about possible modes of operation of the hypothesized linguistic driver. Not only is it assumed that there is "no appreciable lag" between the observed behaviour and the generator: there is also a premise that the measured details can be submitted to an inverse of the generation process and so cast light on the workings of the original driving system. Assumptions of this kind are implicated in the very measures that are often used in the field (e.g., gaze duration, first-pass reading time, regression rate etc.). The rationale for measuring (and basing inferences upon) the time spent inspecting a word or phrase is that, as a direct consequence of tight linkage, this is assumed to tell us something about how long the *generator* devotes to its linguistic tasks. A saccadic movement away from the word is taken as a signal that the driver has completed its work, and has moved on to conduct new operations. Similarly, measuring regression rates only makes sense within a framework in which it is assumed that the backward-moving departure from a word or phrase is a direct reflection of what the driver is doing.

While there is considerable evidence (e.g., Rayner, 1998) that linguistic properties of fixated words (such as their frequency and predictability from prior context) are quite capable of influencing how long readers dwell on those words (demonstrating relatively tight-linkage, at least at this level), there is also abundant evidence that tight-linkage does not always offer an accurate description of the control regime. Spillover effects (e.g., Rayner & Duffy, 1986) demonstrate that a saccadic move away from a word does not reliably signal that its processing is complete and, similarly, preview effects (e.g., Rayner, 1998) demonstrate that at least some aspects of word processing may begin before the word is fixated. Equally, the demonstration that word-skipping depends upon the frequency of a parafoveally placed word (e.g., Reichle et al., 2003, Section 3.2) indicates that lexical processing can start before the word is centrally fixated.

In the analysis that follows, we consider the continuum between *closely-coupled* and *loosely-coupled* eye-movement control. At the latter end of the scale, the relevant term could be applied to systems in which the transmission system between the driver and eye-behaviour includes connections which are subject to random noise. It might also include systems in which the influence imposed by one driver is moderated by that of some unrelated influence. A critical property of loosely coupled systems is that the unpredictable interface makes it indefensible to draw certain kinds of inference. For example, a fixation shift may convey no information at all about the work of the driver. It follows that, within this kind of framework, a change in gaze duration (say) would not warrant drawing any conclusions about linguistic operations.

In examining syntactically generated regressions later in the paper, we highlight examples of regression-control models premised upon the two kinds of interface. As a representative of tightly-coupled accounts featuring purely *linguistic supervision* of regressions we refer extensively to the classic work of Frazier and Rayner (1982). To account for the patterns of regressive eye-movements in a syntactic ambiguity resolution study, they proposed that readers make use of a *Selective Reanalysis* strategy. This enables the linguistic processor to exploit "whatever information it has available about the type of error it has committed to guide its reanalysis attempts" (Frazier & Rayner, 1982: Abstract). Based on this information, the eyes are sent "directly to the ambiguous phrase ...(i.e., the region containing the information that would permit the parser to locate the source of its error)" (p. 188). Regressions of this kind might be described as being both *Linguistically Guided* and *Destination-linked*. The eye-movements are assumed to be purposeful and directed, with the precise landing sites being governed by what might be termed the "linguistic topography" on the page or screen of the sought-for material. Inasmuch as the landing sites are assumed to be under the direct control of the Reanalysis system, we take the Selective Reanalysis system to be one based on tightly-coupled eye-movement control.

As a loosely-coupled foil for Selective Reanalysis we propose a radically different system of regression control. Instead of assuming that the linguistic system sends the eyes to a particular prior location, we posit that its repertoire of interventions is restricted to (i) sanctioning (or, alternatively, vetoing) the implementation of processes that culminate in launching saccades to new (as-yet-unread) material and (ii) triggering a regression to an unspecified earlier word. At any point when the linguistic processor is up to date with its work (i.e., whenever the backlog of its work is below a threshold level), it will signal its readiness to proceed. We assume that the eye control system refers to this status flag in programming any new movement. Given approval to move on, the system proceeds to prepare this movement—with this preparation perhaps even being entirely autonomous (i.e., formulated independently of the syntactic processor). The smooth processing associated with cases of this kind can be contrasted with instances in which the eye-controller is in receipt of a (low-level) signal to move on but finds that the linguistic processor has accumulated a backlog of work and therefore does not show its ready-to-proceed flag. Here we assume that the planned progressive movement is inhibited (perhaps using a cognitive progression-inhibition mechanism along the lines of that proposed by Yang, 2006; p. 59). As a consequence, the eye-controller is induced to programme a saccade that causes the eyes to mark time (and so allow the linguistic processor to catch up with its backlog of work). At this point we assume that information harvesting is scaled back, and perhaps shut down completely as the visual extraction system enters an "idling" mode. Further fixations of the same kind are programmed until the linguistic processor eventually sanctions an advance to new material. For these non-progressing events, the precise location of the fixations is held to be immaterial. However, based on the crude premise that there is a general influence of economy-of-effort and that short saccades are easier to execute than longer ones, we assume that the most prevalent landing sites will be those reached using short saccades (i.e., positions that are very close to the launch site). Within a system of this kind, the linguistic processor is assumed to have a bare minimum of direct control over eye-movements. In most circumstances it merely gives the go-ahead (or veto) when an unrelated system (perhaps the lexical processor) plans a move. On a small minority of trials—associated with the very highest backlog levels—the linguistic processor might also trigger a regression to buy time for the parser to catch up with a backlog of work. In short, the form of eye-movement control is assumed to be extremely loosely-coupled. With this delegated form of control, the landing sites of regressions are assumed not to be linguistically supervised. Instead we posit a more automated form of control, based primarily on the physical and spatial properties of the display. The principles of control for such regressions might be described as being *proximity-based* or *provenance-based*. Within a system of this kind, the purpose of regressive fixations (and, indeed, re-fixations on the same word) is not to refresh the evidence but merely a delaying tactic used to provide "time out" for as-yet-incomplete parsing operations. In view of this *raison d'être*, we refer to this form of control as a *Time Out* generating system, and we use the expression *Time Out hypothesis* to capture the claim that here the function of the system is nothing more than that of postponing new input. We note in passing that a system of this kind is not easily characterized in terms of Lewis (1993) distinction between overt and covert processing. The control process could be described as being *overt* in the sense that processing difficulties could well trigger a distinctive pattern of regressions (characterized by short-range local saccades). However, it could equally be classified as being *covert* on the grounds that any concurrent linguistic work would have to be based on stored information and therefore entirely decoupled from the information represented at the fixation co-ordinates.

In the rest of the paper we evaluate the explanatory potential of these two broad classes of regression control. Of course, systems for delegated and tight control need not be mutually exclusive. It is perfectly possible for one sub-population of regressive movements to be generated by one kind of mechanism and the rest by another. As we review the evidence, we seek to distinguish between three broad classes of account: (1) Systems in which regression control is exclusively tightly coupled with every fixation being clamped onto some linguistic operation (as exemplified by Frazier and Rayner's Selective Reanalysis hypothesis); (2) Control regimes in which the targeting of regressive fixations is entirely decoupled from on-going linguistic processing and (3) Hybrid accounts in which the distribution of landing sites is partly determined by linguistic supervision and partly by either delegated or competing drivers. At various points, we also consider the possibility that certain specifically defined sub-populations of regressions operate according to one set of principles while others are subject to different regimes of control.

In assessing the strength of support for these different accounts, we focus initially on the form of control used to direct the *first* regressive saccade out of a disambiguation region located in a *non-final* region of a sentence. At a later point (in Experiment 2) we track the landing sites of fixations occurring at later points in the regression scan-path. As spelt out above, our reason for restricting our analyses to disambiguation regions prior to the end of the sentence is that we treat this work as a step toward refining theories of *syntactic* processing, with other aspects of sentence processing having much lower priority. There is long-established evidence that final regions of sentences are associated with what have been termed sentence "wrap-up" effects (Just & Carpenter, 1980; Mitchell & Green, 1978; Rayner, Juhasz, Ashby, & Clifton, 2003), arguably implicating high-level discourse processing, coherence checking and other as yet poorly specified operations. There is no guarantee that the forms of eye-control that apply during such wrap-up operations are the same as those relevant to work on parsing. Indeed, in their original study Frazier and Rayner (1982) reported observing eye movements that showed qualitatively different patterns depending on whether participants read through to the end of the sentence or not. In one commonly occurring pattern they noted that the "eye movements generally continued in a forward direction through the sentence. Upon reading the end of the sentence, the subject then made a long regression to the beginning of the sentence and reread the sentence." (Frazier & Rayner, 1982, p. 196). They contrasted this prototypic pattern with a second one in which there were either unusually long fixations on the disambiguation word or, alternatively, sequences in which the "reader fixated on the disambiguation region for an average amount of time and then immediately made a regression back to the ambiguous region of the sentence." (Frazier & Rayner, 1982, p. 197). Given the complex variety of operations that may occur at the ends of sentences, we have taken a tactical decision to prioritize the examination of regressions launched from a region that is unlikely to be contaminated by wrap-up effects.

## Existing evidence for linguistic guidance

Prior work provides surprisingly little evidence that can be used to distinguish between destination-linked and provenance-based accounts of regression targeting (and therefore between the corresponding closely supervised and delegated forms of control). There are some grounds for maintaining that relatively closely coupled linguistic guidance can feature in certain tasks *other* than conventional reading and, within reading, in situations calling for the identification of an antecedent for an anaphor. For example, using a word-matching task, Kennedy and Murray (1987) demonstrated that regressions are frequently sent directly back to a target word. In this particular study, subjects first read a sentence and then a test word located to the right of the target sentence. The task was to say whether the word had appeared at any point in the sentence, and this is reported as having generated very precisely targeted regressions. More recent work using broadly similar target-seeking paradigms have yielded some evidence of spatial selectivity but often without the same degree of precision reported in the original Kennedy/Murray study (Inhoff & Weger, 2005, Experiments 1 and 2; Weger & Inhoff, 2007). In an anaphor resolution study Murray and Kennedy (1988) showed that readers made more large regressive movements to remote antecedents appearing approximately 10 words prior to the pronoun than they did to closer antecedents located at about half that distance (for more recent elaboration of this work see Kennedy, Brooks, Flynn, & Prophet, 2003). In related work using similar anaphoric returns to an antecedent, Inhoff & Weger, 2005, Experiments 3 and 4) found evidence of regressions to the *vicinity* of the target word, but again without great precision—to the extent that further corrective movements were required to reach the target itself. Setting aside uncertainties about the exact precision of the regressive movements, these studies clearly demonstrate that there are circumstances in which the choice of regressive landing sites can be influenced by the linguistic features of a target word. Some form of linguistic supervision is clearly implicated here. The central question for our present purposes, though, is whether a similar form of supervision *also* plays a role in determining the landing sites of regressions launched from a word that marks the end of a temporary structural ambiguity. If it does, then this would provide grounds for ruling out purely provenance-based delegated accounts of control, leaving a further distinction to be drawn between pure linguistic supervision and hybrid models of regression control. To explore these issues it is crucial to go beyond word-matching and co-indexing experiments and focus instead on testing the properties of syntactically-generated regressions.

As already noted, Frazier and Rayner (1982) reported that the "dominant type of regression" in their study consisted of saccades going directly back to the ambiguous region of text. Expressed this way, their summary of the evidence gives the impression that the eyes were being sent back to a particular location in the prior text (exactly in the spirit of their Selective Reanalysis hypothesis). However, since the launch-site in their study (i.e., the disambiguating word marking the end of a temporarily ambiguous clause) was by definition located immediately to the right of the ambiguous region, a return to this part of the sentence can equally well be characterized in Time Out terms as "making the smallest possible regression". Thus it is at least conceivable that the participants fixated on the ambiguous region not because of its syntactic properties (i.e., the ambiguity of the material or its potential as a source of information to guide reanalysis) but simply because it was the portion of the text that was physically closest to the launch-site and therefore the most likely to be programmed by a proximity based system. In effect, this ambiguity of interpretation makes it impossible to use the Frazier and Rayner (1982) data to adjudicate between accounts based on close linguistic supervision and those based on delegated control. Returns to the "ambiguous region" may well be expected to dominate within both of the conceptual frameworks under examination. For many years there has been little investment in further exploration of these issues. However, Meseguer et al. (2002) have since reported what they took to be further evidence of selective (i.e., linguistically supervised) regressions from the ends of disambiguated sentences to an earlier point in the sentence (see Braze, Shankweiler, Ni, & Palumbo, 2002, for other broadly similar findings). Unfortunately, in these studies the differentially targeted regressions were those launched from the very last region of the sentence. As spelt out above, regressions launched from the ends of sentences could well be under the command of wrap-up processes and may tell us little or nothing about syntactic effects or other operations that occur at earlier points in the sentence. To avoid dispute about the form of control in force for syntactically triggered regressions, it is important to build upon evidence that is uncontaminated by influences from unrelated operations.

Most other published studies fail to report regression data in sufficient detail to discriminate between the two different types of account of regression control. To explore this matter further, we now report two new experimental studies. Both make use of the classic pre-posed adverbial subject-object ambiguity discussed and analysed by Frazier (1978) and later used by Frazier and Rayner (1982). In this ambiguity (illustrated in the much quoted sentence (1) below) a noun phrase is temporarily ambiguous between being the object of a preceding subordinate verb and the subject of a later matrix verb.

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

Materials of this kind have been widely investigated in the literature. In the course of this work they have been used to provide an evidence base for disparate variety of theoretical issues from the reanalysis phenomena of central interest here (e.g., Ferreira & Henderson, 1991; Ferreira & Henderson, 1998; Frazier & Rayner, 1982; Warner & Glass, 1987), to questions about whether ultimately incorrect analyses are eventually deleted (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001) and debates about the use of verb-subcategorization information in parsing (e.g., Adams, Clifton, & Mitchell, 1998; Mitchell, 1987; Staub, 2007; Traxler, 2005; van Gompel et al., 2001). More specifically, the materials we used were modelled on sentences used by Ferreira and Henderson (1991), Ferreira and Henderson (1993), and in later studies, to extend the length of the region of ambiguity. It was a deliberate decision to base the present set of tests on a form of ambiguity that has been intensively studied in prior research. By returning to well-worked territory, we aim to build upon accumulated insights concerning the linguistic operations in play at different points in the sentence and to use this background information to focus attention on the interface between these operations and eye-control.

## Experiment 1

In a seemingly trivial but theoretically driven departure from previous work, our first study systematically manipulates the physical position of the disambiguating word (*italicised* below), as illustrated in (2a,b):

| Word 1 | Word 2 | Word 3 | Word 4 | Word 5 | Word 6 | Word 7 | Word 8 | Word 9 | Word 10 | Word 11 | Word 12 |

(2a) While those men hunted(,) the moose that was sturdy and nimble *hurried*

| Region 13 |

into the woods and took cover.

| Word 1 | Word 2 | Word 3 | Word 4 | Word 5 | Word 6 | Word 7 | Word 8 | Word 9 | Word 10 | Word 11 |

(2b) While those men hunted(,) the moose that was sturdy and nimble

| Word 12 | Region 13 |

*hurried* into the woods and took cover.

Across a counterbalanced study, participants are asked to read sentences that appear in one of four different forms: two being temporarily ambiguous as in Examples (2a,b) presented without a comma, and two being unambiguous controls for which the normal punctuation is included. The sequence of words in the sentence was unchanged across all four conditions, and in the remainder of the paper the word numbers (assigned above) are used to refer to specific parts of the sentence. In each experimental sentence, the disambiguating word (here "*hurried*") was Word 12.

The main manipulation was that in half of the sentences—exemplified here by (2a)—the line-break occurred *after* the disambiguation word (i.e., after Word 12: "*hurried*"), whereas in the other half the line-break was introduced *before* this word (as in (2b). Assuming there is an increased incidence of regressions from the disambiguation word in the unpunctuated conditions (reflecting a build-up of linguistic work at this point), the main focus of interest will be the landing sites of the resulting regressions. Note that, by design, the regression region is identical in the two alternative text formats. (By "regression region" we mean either the string of words (Words 1–11) or the region of the display that can act as the target for regressive saccades: in each condition these materials occupy exactly the same pixel positions on the screen). If regressions are under tight linguistic control with landing-site choice being entirely *destination-driven*, then the spatial distribution of these landing sites should be totally unaffected by the position of the line-break. In both layouts the return site should be selected on some linguistic basis and—assuming high-precision control—the two destination distributions should be entirely congruent with one another. In contrast, in more loosely-coupled delegated forms of control offering scope for low-level *provenance* or *proximity* effects, regressions might be expected to move to the closest part of the regression region in each case. Thus, in (2a) they might land on Words 10 or 11, whereas in (2b) the points most easily reached might be in the locations of Words 1, 2 or perhaps 3 (as these appear in a location immediately above the launch site). Accordingly, accounts of this kind therefore predict *differing* patterns of regression landing sites in the two line-break conditions. Evidence of such differences would clearly be incompatible with totally linguistically-driven accounts, but would leave open the possibility that landing distributions are determined by the joint action of both forms of control.

Our data analyses make use of a range of different empirical measures. Because the Time Out and Selective Reanalysis hypotheses differ in the scope of their claims, different analyses are appropriate in each case. Selective Reanalysis is taken to be associated specifically with revisions triggered by higher level linguistic operations (e.g., syntactic processing, but possibly also anaphoric reference and discourse processing). There has never been any claim that its remit extends to *all* regressions. Because of this, the measures best equipped to highlight its properties are those that titrate out the contribution of lower level (e.g., lexical) regressions. In the current experiments this is best achieved by examining the *change* in regression patterns in the unpunctuated relative to the punctuated condition. At substantial distances from the comma, the two displays should be locally identical and low-level regressions should therefore be equivalent. In these circumstances, any shifts in the patterns can reasonably be attributed to the effects of higher-level processes. In contrast with this, the Time Out framework currently draws no distinction between time-marking periods triggered by low-level and high-level operations. As a first step in exploring the potential of the proposal, we assume that all regressions reflect the fact that the eyes have temporarily been placed on stand-by while some unspecified cognitive operation catches up with its backlog of work. Given this, there is no particular rationale within this framework for focusing attention on the way regression patterns *change* with punctuation. By hypothesis, the regressions from the punctuated and unpunctuated conditions do not come from different sub-populations, and so there is no reason to highlight punctuation-based effects. Within the Time Out framework, therefore, the regression patterns *per se* (rather than shifts across conditions) can be viewed as throwing light on the machinery controlling the system. Because of these different requirements, the analyses that follow include treatments of both punctuation-contingent data (for the benefit of Selective Reanalysis hypothesis) and on data that places little emphasis on punctuation distinctions (for Time Out accounts).

## Method

### Participants

Twenty-eight volunteers participated in this study. All were students or employed at the University of Exeter and aged between 18 and 38. Six were male and 22 were female. They were either paid £3–£6 or granted equivalent course credit for their participation if they were first year Psychology students. In total, eight of the participants were paid in cash. The whole experiment normally took up to 30–50 min depending on the reading speed of different participants.

### Materials

The study used the "Early" forms of the 24 sentences listed in the Appendix. (In this description, "Early" and "Late" refer to the positions of the area of potential structural misanalysis. In the present experiments, difficulties might creep in after Word 4, whereas in Experiment 2, below, the region of ambiguity can also start "Late"—after Word 9). Across the experiment as a whole, each sentence was presented in four different forms, always using precisely the same string of individual words. In two of the forms, there was a pre-disambiguation Line-break immediately after Word 11 and before the disambiguating word (which—as already indicated—was always Word 12). In the remaining two Line-break conditions Word 12 was kept on Line 1, with all the remaining words of the sentence being shifted to Line 2. This Line-break or layout manipulation was crossed with a manipulation of punctuation. In two of the forms, the sentence appeared *without* a comma (but with a conventional period at the end). In the other two, a comma was inserted immediately after Word 4.

In each word position, word-length was matched within a range of two characters. For example, Word 12 was always 7, 8 or 9 characters long (mean: 7.9 characters). In certain cases there was no variation at all: Words 5 and 7 were always exactly three characters long. Counting backwards from the end of Word 11, each preceding space between words fell in a very similar location across all 24 sentences. For example, the gap between Words 10 and 11 falls 5–7 characters back, the one before that 9–11 characters back (with the extra comprising the word "and" plus a space), and so on. The beginning of the sentence was always 59–61 characters (including spaces) to the left of this reference point. This was designed to ensure that the layout of Words 1–11 (the "regression region" for any regressions launched from Word 12) was almost entirely uniform. To preserve this feature we used a typeface (font: Bitstream Vera Sans Mono Bold; font size 19) with non-proportional (monospaced) characters with intervals of 11 horizontal pixels per character (with actual characters occupying 8–10 pixels depend on width: e.g., "j" or "w" at the two extremes). Note that because the comma also occupied a standard 11-pixel slot, in the punctuated condition Words 1–4 were shifted 11 pixels further from the reference point than they were in the unpunctuated sentences. Relative to the reference point, there was no shift in for Words 5 through 11. Characters like "o" and "n" were 11 pixels in height (x-height), ascenders extended another 4 pixels above the baseline (making the full ascent height 15 pixels), and descenders dropped 4 pixels below the baseline. The line spacing between Line 1 and Line 2 was set to 60 pixels with the starting point sent to co-ordinates (10,200). This line separation is the equivalent of at least double line-spacing. Our decision not to use close line-spacing was intended to minimize the risk of misclassifying return sweeps from Line 1 to the beginning of Line 2 being as regressions to the beginning of Line 1 (and *vice versa*).

In addition to the test sentences, there were 24 two-line foil sentences also starting with pre-posed adverbial clauses but with the noun-phrase ambiguity always resolved as the object of the subordinate verb (i.e., the opposite resolution to that for the experimental sentences). For half of these sentences the end of the pre-posed clause was marked by a comma. Finally, there were 80 additional two-line filler sentences covering a wide range of different structures. As before, half of these included commas at felicitous points.

A simple yes/no comprehension question was formulated for 6 of the test sentences and 24 of the remaining sentences. Participants were expected to choose one correct answer after they finished reading each of the sentences.

The experimental session itself was preceded by a practice session comprising five sentences together with two comprehension questions.

The test materials were counterbalanced such that each participant read every one of the 24 sentences, with exactly six exemplars formatted for each of the $2 \times 2$ Comma $\times$ Line-break conditions. The materials were rotated such that across the experiment as a whole each sentence appeared equally often in each of its four formats. Within an individual experimental session, sentences themselves were presented in a pseudo-random order.

*Apparatus*

The apparatus was an EyeLink® II eye tracker developed by SR Research Ltd., connected to two Dell computers. The eye tracker has a head mounted system with two miniature cameras mounted on a comfortable padded helmet and an extra camera in the middle of the helmet to determine the central position of the head. The two eye cameras allowed binocular eye tracking with built-in illuminators in each of them. The screen was set to the resolution of 800*600 pixels and the top left of the first letter of each sentence was located at screen coordinate (6, 218). The experimenter's computer was equipped with the EyeLink® II set-up and control programme so that all the calibration and validation could be controlled through this screen.

*Procedure*

On entering the test cubicle, each participant was asked to put on the eye-tracking helmet. One of the EyeLink cameras was directed at the participant's right pupil. At the beginning of the session approximately 10 min was set aside for tracker calibration. The experiment proper was started only after calibration and validation was classified as being "Good" within the standard EyeLink system. Once this had been achieved, participants were presented with five practice sentences. In each case the display was initiated when the participant pressed a button to advance to the next trial. A further press triggered the display of a question (where this was included). Comprehension questions were answered by pressing game controller buttons marked either "Yes" or "No". Calibration and validation were freshly adjusted between the presentations of successive sentences. A fixation point was displayed at pixel (10,200) to mark the starting point of the sentence. Participants were instructed to press the Advance button once they had focussed on this dot. Provided the tracker returned the same coordinates at this point (modulo small prespecified tolerances), the dot display was removed and replaced immediately by the display of the new sentence. In cases where the discrepancy threshold was exceeded, there followed an automatic recalibration prior to the display of the new text. Participants were invited to ask questions during the practice session. On its completion, the experimenter left the test cubicle, allowing the participant to work their way through the full experimental session.

*Results*

Participants made errors on just 7.5% of comprehension questions overall, indicating that they paid reasonable attention to the content.

Standard eye-tracking measures were calculated for each of the first twelve words plus a region combining Words 13 and 14 (referred to throughout as Region 13). For each word, an "Interest Area" was defined as comprising a rectangle bisecting the space on either side of the word, and extending 30 pixels above the height of ascender characters and 30 pixels below the baseline (with no overlap between lines). For each of the measures below (other than Total reading time) a trial was treated as yielding missing data if the very first fixation on a word or region itself followed a sequence of fixations that entailed first *skipping* the word/region under scrutiny and then fixating on a later word. The exclusion of data from 'backward approaches' from these measures serves to guarantee that the calculated metrics are not contaminated by the effects of regressions from later material. The different eye-tracking measures were defined as follows: *First fixation time*: Duration of the very first fixation falling within a defined Interest Area (IA) without having been preceded by fixations in IAs with a higher index; *First-pass reading time*: Sum of the First Fixation time together with that of all other fixations falling within a defined IA up until the point at which a saccade first crosses the IA boundary in any direction. (Following the first such exit, no further return visits have their times included in the tally). In the literature, this measure is also referred to as *gaze duration* in cases where the IA consists of a single word; *Percentage of first-pass regressions*: the percentage of all first-pass episodes (as defined above) which are followed by a regression rather than a progression on to later text. (*Regressions* were defined for the present analysis as saccades departing from the current IA and landing in an IA with a lower index (i.e., the interest area associated with a previous word). *Progressions* were equivalently defined as departing saccades that land in an IA with a higher index.); *Regression-path duration*: the total of all fixations included within the first pass episode for an IA, plus the sequence of all fixation times following a regression up until the point at which there is a saccade despatched to an IA with a higher index (for words in locations other than the end of the line this normally involves crossing the right-hand boundary of the IA for the current word, but for words near the end of the line saccades sweeping to the beginning of the new line my not cross this boundary). In the literature, this measure has been given a variety of labels including *regression path time, Cumulative Region Regression Time* (CRRT) and *go-past time*. One further traditional measure used was *Total Reading time*: the total of all the fixations that fall within a defined Interest Area at any point in the course of reading the sentence. For any trial on which there were no visits at all, the value recorded was 0 ms. Total reading time was the only timing measure that could return a zero value. Finally, for this study we generated a new measure, which we term the *Regression Signature*. This is the distribution of word landing sites—or more technically IAs - for the set of first regressive saccades launched from any IA (i.e., word or word-combination) under scrutiny. To facilitate comparisons across conditions (e.g., to compare the Word 12 Regression Signatures for punctuated and unpunctuated sentences) the frequency of each participant's returns to each potential IA landing site is expressed as a proportion of the total number of first regressions the participant made from the same word/IA (and in the same condition). This makes it possible to compare signature shapes over conditions with very different regression frequencies. For purposes of data-plotting and statistical comparison, regression data for an individual participant are only retained in cases where there are genuine regressions (as defined above) in each of the conditions being compared. That is, if a participant has no regressions in one particular condition, then *all* of that participant's data are routinely removed from the analysis in question. In defining the regression corpus from Word 12 in addition to the regressions automatically identified by the EyeLink software, we added 6 "close misses" that fell slightly higher than the standard 30-pixel vertical limit. In these cases the horizontal IA boundaries were used to determine the closest word and therefore to specify a landing site. This criterion relaxation was used to increase by a few events the size of the corpus of first regressions to be used for detailed analysis. In all other analyses, where sample-size is less of an issue, saccades falling beyond the vertical limits were treated as missing data. Finally, as a statistical reporting convention we follow the recommendations of Masson and Loftus (2003) and report 95% Confidence Intervals (CIs) for each of our effects. In every case these are based on the error estimates from the analysis by participants and are always calculated individually for the effect in question (i.e., we avoid the use of CI bands based on pooled error estimates).

The ultimate focus of interest in this study is intended to be the various regression signatures launched from Word 12 (the disambiguating word) across the different experimental conditions. However these regression signatures are only worth examining in detail if we can first establish that Word 12 is the focus of specialised syntactic processing associated with ambiguity resolution We therefore carried out a series of preliminary tests designed first to scrutinize the characteristics of processing across the first eleven words and secondly to confirm that the reading time for Word 12 showed the expected increase in the unpunctuated than in the punctuated (and therefore unambiguous) condition. To test for these patterns, for each of the first eleven words we extracted each of the standard eye-tracking measures: namely, first fixation time, first-pass time, percentage of first-pass regressions, regression-path time and total reading time. For each measure and separately for each word, the data were submitted to two-way ANOVAs with repeated measures for both Comma (present or absent) and Line-break (line-break between Words 12 and 13, leaving the disambiguating word at end of Line 1 vs. line-break between Word 11 and 12, shifting the disambiguating word to the beginning of Line 2). In several cases, one or more of the participants (or sentences) had entirely missing data resulting from the target word having been skipped in all cases representing a particular sub-condition. In all such instances, all data from the participant (sentence) were excluded from the analysis in question, resulting in some fluctuation in the degrees of freedom in the analyses reported below. The analyses themselves were comprehensive but, given the scene-setting nature of these initial observations, we merely summarize the findings for Words 1-11 and for Word 12 we restrict our reporting to effects that were reliable both on the participants' and the materials analyses.

Across the first eleven words, five of thirteen reliable effects were total reading time main effects that were arguably generated by increased rates of returns from later regions, thus providing little information about the initial processing of the material. For example at Word 2 the total reading time was higher in the unpunctuated than the punctuated condition despite the fact that this word was centered 12–13 characters to the left of the manipulated comma position. Similar total reading time punctuation effects emerged at Words 4, 5 and 7. At Word 11, total reading time was markedly increased in the Late linebreak condition where this word was flanked by Word 12. This figure was boosted by substantially more returns from later material than it was when the following word was shifted to a new line—which apparently inhibited such returns. In the case of eye-tracking measures that reflect more immediate processing as opposed to revisits during regressions from later material, there was evidence of some processing turbulence in the vicinity of the comma. At Word 4 there was an increased rate of regressions when the comma was *present* together with a corresponding increase in regression path duration. This suggests that the introduction of the comma precipitates processing that is different from that taking place in its absence. At Word 11 there were higher regression rates and shorter first-pass reading times in the early linebreak condition (i.e., when the break occurred immediately after Word 11 itself). Specifically there was a 16.7% regression rate when the linebreak was between Words 12 and 13, compared with a rate of 26.2% in the early break condition (between 11 and 12), CI ± 6.9%, $F_1(1,24) = 8.41$, $p < .05$; $F_2(1,23) = 10.54$, $p < .01$; $\text{Min}F'(1,46) = 4.68$, $p < .05$). The corresponding first-pass results were: Late linebreak—307 ms.; Early linebreak—257 ms., CI ± 37 ms, $F_1(1,24) = 7.90$, $p < .01$; $F_2(1,23) = 7.65$, $p < .05$; $\text{Min}F'(1,46) = 3.89$, $.05 < p < .1$).

Focussing next on events taking place in the disambiguation region, we now turn our attention to processing effects associated with Word 12 and the two words following that (here termed Region 13, as already indicated above). Table 1 provides a summary of the means, F-ratios, and confidence intervals for each of the standard eye-tracking measures in each of these locations. These data show evidence of a Comma main effect at the disambiguation word itself (Word 12) in each of the regression-based measures (regression path duration and regression rate). This effect also spilt over and showed up in the same two measures in Region 13. These results show clear evidence of increased regression activity at these locations in the unpunctuated relative to the punctuated conditions. This is consistent with the expectation that there would be processing difficulty in cases where Word 12 resolves a temporary structural ambiguity. Powerful punctuation effects were also evident in the total reading time measures in both locations. The results also showed marked Linebreak effects in the regression rate for both regions and in the total reading time measure for Word 12. Processing associated with Word 12 caused a higher rate of regressions when this word appeared at the end of the line. However the total reading time was *reduced* in this condition compared to the situation in which Word 12 appears at the beginning of Line 2. In Region 13 there was a much *higher* rate of regressions in the Early than in the Late linebreak condition (the exact opposite of the Word 12 pattern), and there was also a highly reliable first pass time effect with markedly longer times spent in the region in the Late linebreak condition. Detailed interpretations of these patterns of results will be explored below, but the main findings can readily be understood on the premise that there is a marked inhibition about programming regressions that involve making eye-movements to the line above as opposed to leftward saccades that keep to the same line of text. At Word 12 this would act to reduce regression rates in the Early linebreak condition (in which the word appears in the first location on Line 2). For Region 13 it would have the opposite effect: namely reducing regression rates in the *Late* linebreak condition, in which it is Word 13 that now occupies the first position in Line 2.

Having established that there are fairly standard disambiguation effects evident in the data (in addition to the much more complex Linebreak effects), we can now return to the primary focus of interest for this experiment, namely examining whether there is any evidence that the landing sites for regressive saccades are modulated under the different conditions of the study.

Using conventions set out above, various corpora of regressions were identified for detailed analysis. The first was the corpus of all regressions launched from Word 12 (the disambiguating word). This consisted of 87 regressions in all: 45 in the Unpunctuated condition with Word 12 appearing at the end of line 1; 27 in the Comma control (i.e., unpunctuated) condition (again for the Line 1 location); 9 in the Unpunctuated condition with Word 12 appearing at the beginning of Line 2 and 6 in the (punctuated) control for the Line 2 location. The second was a corpus of regressions launched from Region 13 and returning to at least one of the first eleven words. The analyses reported above indicate that there is at least some degree of processing spillover after Word 12, and this second set of regressions can be viewed as events triggered by slightly delayed disambiguation operations. There were 42 such regressions—with 12, 0, 23, and 7 in the individual experimental conditions (in the order given immediately above). A third corpus was constructed by combining these two sets but then removing from the tally any previously included regressions from Region 13 which occurred for the same participant and trial as one of the regressions launched from Word 12. This comprised a total of 123 regressions, respectively, with 53, 27, 30 and 13 in each of the four sub-conditions. This particular corpus covers all instances of the first regression scan-path launched from either Word 12 or Region 13 and extending back into the region covered by Words 1-11 (i.e., the *regression region*).

The main purpose of the present analysis is to examine experimental changes in the distribution of *landing sites* of the first saccade entering the regression region (i.e., alterations in the regression signatures across the different conditions). Our original intention had been to analyse and compare these signatures across the $2 \times 2$ main conditions both for regressions launched from Word 12 and for those originating in Region 13. However, this proved impossible given the paucity of regressions from launch-sites at the beginning of Line 2. At Word 12, just one of the 28 participants had regressions in all four of the Comma × Linebreak conditions. Thus, the extensive amount of missing data ruled out the use of fully orthogonal $2 \times 2$ comparisons. To address this, we restricted our analysis to two pairs of conditions, setting aside entirely the very sparse data from the last of the four conditions. To ameliorate data-loss, we also concentrated our analysis on the pooled corpus of regressions from Word 12 and Region 13 (i.e., Corpus 3, as defined above).

The first analysis was restricted to the two unpunctuated conditions and compared the regression signatures

**Table 1**
Mean reading times and regression rates across the Comma and No Comma versions of sentences with Early and Late linebreaks

| | Mean eye-tracking measure for: No_Comma/Early break; No_Comma/Late break; Comma/Early break; Comma/Late break | *F*-ratio For (in order): Punctuation Linebreak; Punctuation × Linebreak | | | | 95% Confidence interval |
|---|---|---|---|---|---|---|
| *Word 12* | | | | | | |
| First Fixation (in ms.) | 275 | **P** | $F_1(1,22) = 1.79$ | $F_2(1,23) = 3.76$ | | 24 |
| | 216 | **L** | $F_1(1,22) = 1.74$ | $F_2(1,23) = 7.04^{*}$ | | 55 |
| | 235 | **P × L** | $F_1(1,22) = 3.54$ | $F_2(1,23) = 3.22$ | | 26 |
| | 225 | | | | | |
| First Pass time (in ms.) | 333 | **P** | $F_1(1,22) = 2.60$ | $F_2(1,23) = 2.75$ | | 37 |
| | 358 | **L** | $F_1(1,22) = 2.11$ | $F_2(1,23) = 1.62$ | | 56 |
| | 290 | **P × L** | $F_1(1,22) < 1$ | $F_2(1,23) < 1$ | | 32 |
| | 344 | | | | | |
| Regression path duration (in ms.) | 609 | **P** | $F_1(1,22) = 4.89^{*}$ | $F_2(1,23) = 8.24^{**}$ | $\mathrm{Min}F'(1,41) = 3.07$ | 116 |
| | 471 | **L** | $F_1(1,22) = 3.76$ | $F_2(1,23) = 9.75^{**}$ | | 110 |
| | 450 | **P × L** | $F_1(1,22) < 1$ | $F_2(1,23) = 1.99$ | | 95 |
| | 382 | | | | | |
| Total reading time (in ms.) | 411 | **P** | $F_1(1,22) = 22.36^{**}$ | $F_2(1,23) = 23.02^{**}$ | $\mathrm{Min}F'(1,44) = 11.34^{**}$ | 67 |
| | 514 | **L** | $F_1(1,22) = 8.74$ | $F_2(1,23) = 10.93$ | | 83 |
| | 267 | **P × L** | $F_1(1,22) < 1$ | $F_2(1,23) < 1$ | | 41 |
| | 379 | | | | | |
| Regression rate (%) | 28.5% | **P** | $F_1(1,22) = 3.87^{*}$ | $F_2(1,23) = 5.49^{*}$ | $\mathrm{Min}F'(1,43) = 2.27$ | 6.4% |
| | 7.3% | **L** | $F_1(1,22) = 16.87^{**}$ | $F_2(1,23) = 36.04^{**}$ | $\mathrm{Min}F'(1,39) = 11.49^{**}$ | 8.6% |
| | 18.3% | **P × L** | $F_1(1,22) = 2.49$ | $F_2(1,23) = 2.49$ | | 5.1% |
| | 5.0% | | | | | |
| *Region 13* | | | | | | |
| First Fixation (in ms.) | 206 | **P** | $F_1(1,24) < 1$ | $F_2(1,23) = 3.28$ | | 15 |
| | 206 | **L** | $F_1(1,24) < 1$ | $F_2(1,23) = 2.38$ | | 27 |
| | 207 | **P × L** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 14 |
| | 196 | | | | | |
| First Pass time (in ms.) | 340 | **P** | $F_1(1,24) = 3.29$ | $F_2(1,23) = 2.75$ | | 22 |
| | 264 | **L** | $F_1(1,24) = 74.37^{**}$ | $F_2(1,23) = 69.54^{**}$ | $\mathrm{Min}F'(1,46) = 35.94^{**}$ | 20 |
| | 328 | **P × L** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 24 |
| | 237 | | | | | |
| Regression path duration (in ms.) | 539 | **P** | $F_1(1,24) = 5.73^{*}$ | $F_2(1,23) = 19.36^{**}$ | $\mathrm{Min}F'(1,36) = 4.42^{*}$ | 266 |
| | 798 | **L** | $F_1(1,24) = 2.19$ | $F_2(1,23) = 4.07$ | | 182 |
| | 343 | **P × L** | $F_1(1,24) = 2.19$ | $F_2(1,23) = 2.59$ | | 174 |
| | 368 | | | | | |
| Total reading time (in ms.) | 476 | **P** | $F_1(1,24) = 18.68^{**}$ | $F_2(1,23) = 12.92^{**}$ | $\mathrm{Min}F'(1,45) = 7.64^{**}$ | 63 |
| | 439 | **L** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 53 |
| | 323 | **P × L** | $F_1(1,24) = 2.67$ | $F_2(1,23) < 1$ | | 37 |
| | 343 | | | | | |
| Regression rate (%) | 15.8% | **P** | $F_1(1,24) = 11.36^{**}$ | $F_2(1,23) = 5.84^{*}$ | $\mathrm{Min}F'(1,42) = 3.86$ | 8.2% |
| | 36.9% | **L** | $F_1(1,24) = 18.60^{**}$ | $F_2(1,23) = 107.32^{**}$ | $\mathrm{Min}F'(1,32) = 15.85^{**}$ | 11.3% |
| | 0.0% | **P × L** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 7.3% |
| | 26.0% | | | | | |

Also listed are the F-ratios and 95% confidence intervals for the two main effects and their interaction. The top half of the table gives the results for Word 12 (the disambiguating word) and the remainder covers Region 13 (the post-disambiguation region). A single asterisk indicates $p < .05$, and a pair indicates $p < .01$.

for the Early and Late Linebreak conditions (see Fig. 1). The individual 11-argument probability vectors underlying these Regression signatures were entered into two-way repeated measures participants' and materials ANOVAs with the factors Comma (two levels) and initial Landing Site (11 levels). Because the analyses were restricted to the corpus regressions, and every regression was allocated to exactly one landing site, the sum of the eleven probability values is exactly 1.0 in each of the comma conditions, making it impossible to compute a Comma main effect (the F-ratio being zero divided by zero). The remaining results showed a small main effect of Landing Site ($F_1(10,120) = 2.20$, $p < .05$; $F_2(10,160) = 3.19$, $p < .001$; $\mathrm{Min}F'(10,252) = 1.30$, $p = .23$), providing some indication that the 11 sites differed in the probability with which they were used as initial entry points. There was also a similarly modest Landing Site by Comma interaction: ($F_1(10,120) = 2.39$, $p < .05$; $F_2(10,160) = 2.97$, $p < .01$; $\mathrm{Min}F'(10,263) = 1.32$, $p = .22$). A pairwise comparison showed that of all the revisits to the regression area (Words 1–11) Word 11 was selected more frequently as the first entry point in the Late than in the Early linebreak condition (Late: 44.9% of all first returns; Early: 11.5%, CI ± 22.5%, $F_1(1,12) = 10.40$, $p < .01$; $F_2(1,16) = 9.96$, $p < .01$; Min-

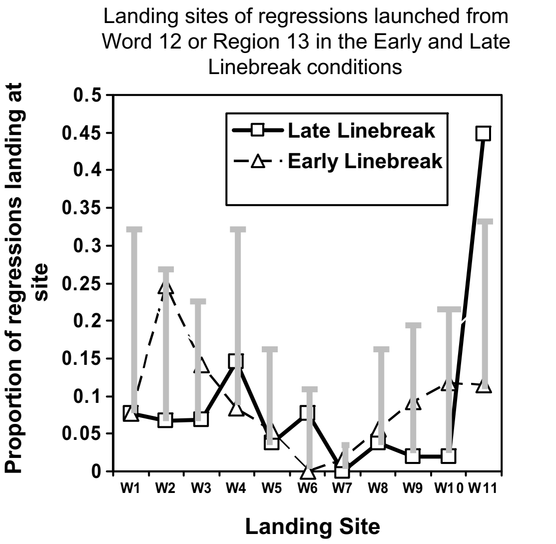


**Fig. 1.** Two Regression Signatures for first regressions launched from Word 12 in its position at the end of Line 1. One signature gives the proportions of regressions sent back to each of the eleven preceding words (labeled W1–W11) for unpunctuated materials (sentences displayed without a Comma). The second gives the equivalent distribution for sentences presented *with* a Comma. Note that this plot does not cover the full corpus of first regressions from Word 12. Data were excluded from the plot for each of the 15 participants who produced no Word 12 regressions from either or both of the punctuation conditions. To avoid cluttering the graph, 95% confidence intervals are given only for the lower of the two data points in each location, and only the upper limit is plotted.

$F'(1,27) = 5.09$, $p < .05$). Comparisons for the other ten words indicated that individually none of them displayed layout-controlled differences that were significant either in the participants' or the materials analysis. (To compensate for the Word 11 effect, clearly Words 1–10 must collectively have attracted fewer initial returns in the Late condition, but the incidence of such returns was too sparse to establish any statistical differences across the first ten words.)

In order to examine the effect of *punctuation* on entry points, a broadly similar set of analyses were run comparing the regression signatures in the punctuated and unpunctuated Late Linebreak conditions (see Fig. 2). In this case there was a substantial main effect of Landing Site ($F_1(10,140) = 42.73$, $p < .001$; $F_2(10,170) = 20.11$, $p < .001$; Min$F'(10,289) = 13.67$, $p < .001$), indicating that the 11 sites differed in their tendencies to attract the first revisit. There was also a robust Landing Site by Comma interaction: ($F_1(10,140) = 6.83$, $p < .001$; $F_2(10,170) = 6.46$, $p < .001$; Min$F'(10,308) = 3.32$, $p < .001$). Word by word pairwise comparisons were conducted to tease apart the details of this interaction. The results showed that Word 11 was selected more frequently as the first entry point in the punctuated than in the unpunctuated condition (With comma: 92.9% of all first returns; Without comma: 45.6%, CI ± 27.4%, $F_1(1,14) = 13.77$, $p < .01$; $F_2(1,17) = 17.44$, $p < .001$; Min$F'(1,27) = 7.695$, $p = .01$). Comparisons for the other ten words indicated that just one (Word 9) showed a hint of a difference—though not one that was reliable by participants (With comma: 0% of all first returns; Without comma: 17.2%, CI ± 19.5%, $F_1(1,14) = 3.60$, $.05 < p < .1$; $F_2(1,17) = 4.75$, $p < .05$).

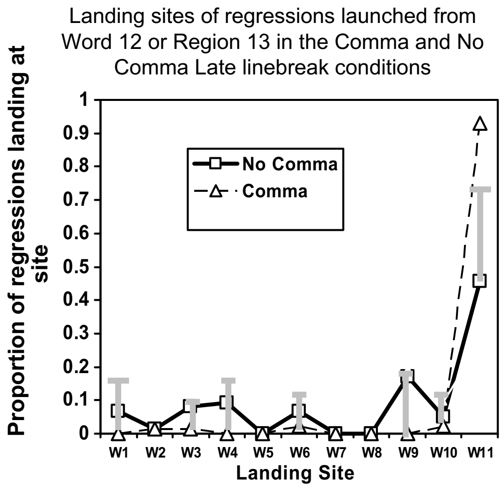


**Fig. 2.** Two Regression Signatures for first regressions launched from Word 12 in its position at the end of Line 1. One signature gives the proportions of regressions sent back to each of the eleven preceding words (labeled W1–W11) for unpunctuated materials (sentences displayed without a Comma). The second gives the equivalent distribution for sentences presented *with* a Comma. Note that this plot does not cover the full corpus of first regressions from Word 12. Data were excluded from the plot for each of the 13 participants who produced no Word 12 regressions from either or both of the punctuation conditions. To avoid cluttering the graph, 95% confidence intervals are given only for the lower of the two data points in each location, and only the upper limit is plotted.

*Discussion*

The experiment was designed primarily to check for evidence of provenance- or proximity-based regression signatures in an ambiguity-resolution study. Preliminary analyses showed—as anticipated—that compared with the punctuated (and therefore presumably disambiguated) materials, the unpunctuated materials showed differential increases in regression rates and in various reading time measures. These patterns clearly have implications for broader accounts of regression supervision. The effects appeared at Word 12 (the disambiguating word) and further spillover effects also emerged in the region comprising the next two words (Region 13). Prior to this there was an unanticipated effect featuring an increased prevalence of regressions from Word 11 when it was located at the end of Line 1 (compared with when it was followed by a further word). There was also evidence of processing difficulty in the vicinity of the comma, when present (e.g., at Word 4). Overall, these details are consistent with the working assumption that, where the comma is included, it can be used as a cue to bias the interpretation of the test sentences in such a way as to favor the interpretation eventually supported toward the end of the sentence. Equally, and in line with a substantial body of prior work, the eye-tracking data suggest that in the absence of punctuation there is initially a bias to interpret the complex noun phrase starting at Word 5 not as the subject of the matrix clause but as the object of the preposed subordinate verb. This leads to processing problems as soon as this second interpretation proves to be untenable (i.e., at Word 12 or a little later). Assuming this much is granted, we take it as a premise that the overall corpus of regressions launched from Word 12 (and also from Region 13) is likely to include substantial subsets triggered either immediately or at a short lag by the heightened demands of the disambiguation process.

The present experiment was designed to determine whether these regressions are influenced by the physical layout of the material (as predicted, for example, by the Time Out hypothesis), or whether their prevalence and spatial characteristics are driven exclusively by linguistic factors (as suggested by a strict interpretation of the Selective Reanalysis hypothesis). In the event, the results showed evidence for both spatial and linguistic influences, perhaps suggesting that a hybrid model might come closest to providing a full account of the data. In support of non-linguistic layout effects there was powerful evidence that the overall probability of making regressions alters as a function of the way in which the material is displayed. For example, in the unpunctuated conditions regressions launched from Word 12 were almost four times as common when this word appeared at the end of Line 1 than in the Early linebreak condition which shifted the word into a position at the beginning of Line 2. This finding suggests that there may be some kind of inhibition of regressions that involve travel to an earlier line. In line with this notion, relative to the Early linebreak condition the rate of regressions from Region 13 was *reduced* in the Late linebreak condition (where this is the material now occupying the slot at the beginning of Line 2). In the Early condition, Word 12 occupies this slot and becomes an attractive destination for the first regressive saccade from Region 13. As in the case of Word 12, regressions are far more common where they can be executed without the need to shift to the previous line. In a further layout effect, this time apparently unconnected with syntactic processing, Word *eleven* featured more regressions when it appeared at the end of Line 1 than when it was juxtaposed on the immediate right by the next word of the sentence. Taken together these findings indicate that layout has a profound effect on regression behavior, and it is clear that there are serious shortcomings in any model of regressions that fails to take account of these phenomena.

Thus far, the results argue emphatically against purely linguistic accounts of regression control (such as unmoderated versions of the Selective Reanalysis hypothesis or any member of the class of models in which regressions are interpreted exclusively as a side-effect of the retrieval of verbal information (e.g., Weger & Inhoff, 2007). However, as they currently stand, the findings do not provide definitive support for proximity-based accounts *per se*. In most conventional formats, the line above is no further away than words to the left, and so proximity accounts offer no basis for favoring same-line regressions. However, that said, more detailed features of the results did provide some support for proximity-based accounts. Specifically, regression signature comparisons between the two linebreak conditions for unpunctuated materials indicated that after due correction for differences in the overall *rates* of regression, the point of entry into the regression area was more likely to be Word 11 when this word was close by (immediately to the left of the disambiguation word) than when it was located at the far end of the line above. Like the earlier evidence that layout has effects on gross regression rates, this finding again underlines the fact that regressions cannot be under exclusively linguistic control. In a situation where word strings were completely matched, the position of first entry into the regression area was demonstrably influenced by the physical provenance of the regression.

If these findings provide clear evidence of layout effects, including some support for proximity effects, it remains the case that they still fall short of ruling out a role for linguistic guidance. Further regression signature comparisons showed that when linebreak was held constant, initial entries into the closest part of the regressions area (i.e., Word 11) were more prevalent with punctuated than with unpunctuated materials. Building upon our working assumptions about the punctuation effects, this suggests that the proximity influences discussed immediately above have a more powerful influence within the particular population of regressions that is less likely to be dominated by disambiguation-linked phenomena (i.e., that exogenous factors are more important for non-syntactic than for syntactically triggered regressions). With syntactically-launched regressions there seems to have been some kind of reduction in pressures that favor proximity-based target selection (i.e., a de-emphasis of forms of guidance that would otherwise encourage the choice of the closest word within the regression region as the entry point for the first return). Of course, under Selective Reanalysis this might occur because a competing (linguistic) driver takes over and directs the saccade to a more distant part of the regression region (e.g., the area in which the initial analysis might have occurred.) However the punctuation-linked shift in entry points falls a long way short of providing unequivocal support for linguistic guidance. An alternative interpretation of the effect might be that that the presence or absence of the comma within the regression region provides a form of physical differentiation which prompts the choice of different landing sites in the two cases. Taking this perspective, the changes in entry point would be seen not as an overt manifestation of linguistic guidance, but merely as yet another layout effect in which saccade destinations are influenced by the physical appearance of the text. Equally, within a Time Out framework it is easy to imagine a system in which the response to an unusually heavy workload is to program not a single local regression but rather to select the characteristics of the first of a string of several successive regressions together intended to buy time to complete the linguistic work in question. On this scenario, it is conceivable that a saccade planned as the first of many may well be sent further back in the text than one programmed to proceed an immediate progression to new material.

This punctuation-based shift in regression landing sites is difficult to account for on a purely physical basis, and it may well be reasonable to treat it as a rather indirect manifestation of the workings of Selective Reanalysis mechanism. Unfortunately, the present findings failed to provide direct evidence for returns to particular points in the sentence. However, supporters of Selective Reanalysis could well object that the present study lacks power, and therefore cannot be presented as offering a robust test of the hypothesis. It is certainly the case that on half of the trials, shifting the disambiguating word to a new line dramatically reduced regression rates, therefore presumably reducing the chances of detecting relatively subtle linguistic guidance effects. Experiment 2 introduces a number of changes designed to increase the sensitivity of measurement.

## Experiment 2

The results of Experiment 1 provided clear evidence that the destinations of regressive saccades are influenced by the text format of the launch site. There was also the suggestion of a possible linguistic guidance effect inferred from the finding that the regression signature for the Line 1 disambiguating word differed across the comma-disambiguated and unpunctuated conditions. However, this inference was compromised by the fact that the presence or absence of the comma itself changed the physical appearance of the text, leaving scope for layout-based explanations of the effect. To rule out counter-explanations of this kind, a compelling demonstration of linguistically guided effects would have to be based on a comparison between prior texts that can be treated as being physically indistinguishable (i.e., with characters and spaces coinciding across the two conditions). In addition, it goes without saying that there also needs to be a rationale for expecting a linguistically guided reanalysis system to send the eyes back to different locations within this region.

To address these various requirements, the approach we adopt here is to use sentences like (3a,b):

| Word 1 | Word 2 | Word 3 | Word 4 | Word 5 | Word 6 | Word 7 | Word 8 | Word 9 | Word 10 | Word 11 | Word 12 |

(3a) While those men **hunted(,) the moose** that was sturdy and nimble *hurried*

| Region 13 |

into the woods and took cover. (Early Misanalysis Area)

| Word 1 | Word 2 | Word 3 | Word 4 | Word 5 | Word 6 | Word 7 | Word 8 | Word 9 | Word 10 | Word 11 | Word 12 |

(3b) One sole hiker spotted that while those men **hunted(,) the moose** *hurried*

| Region 13 |

into the woods and took cover. (Late Misanalysis Area)

The Early Misanalysis Area condition is identical to the same-line condition used in Experiment 1. However, in the Late Misanalysis Area condition the clause boundary (and comma insertion-point) is shifted five words to the right. In both versions of the sentence, the disambiguating word ("hurried" in Example 3b) is preceded in turn by a 5- or 6-letter word, a 3-letter word, 6-letter word and so on. The arrangement of words and spaces is closely congruent across the two conditions. In sentences like (3a), the head of the phrase initially likely to be misparsed appears as the sixth word ("moose", in this particular case). In the version represented by (3b), the corresponding structure appears in the position of Word 11. In both cases the word to which the noun phrase is likely to be misattached ("hunted") appears two words to the left of the ambiguous noun phrase head.

We start by making the crude assumption that any machinery responsible for Selective Reanalysis would send the eyes back to somewhere near the point where things first went wrong (i.e., somewhere in the region *hunted...moose*, where the two units might have been wrongly linked in a premature syntactic commitment). On this premise, for sentences with Early Misanalysis Areas the machinery would be expected to aim for landing sites that are on average five words to the left of what would be appropriate for materials with Late Misanalysis Areas (e.g., Sentence 3b). This 5-word difference in peak target distributions would be expected to show up in the form of suitably modified regression signatures in the two conditions (with a shift in the peak of visiting preference to the left in cases like 3a).

In order to add further precision to these predictions, we draw upon published accounts of the linguistic features of reanalysis in this particular structure (e.g., Ferreira & Henderson, 1998, pp 76-97; Fodor & Inoue, 1998, pp. 114–126; Stevenson, 1998, pp. 355–357). As already indicated, there have been no explicit predictions about how the effects of reanalysis might manifest themselves in the eye-tracking record. We are therefore compelled to draw our own inferences from the different accounts of reanalysis put forward. In describing the forms of reanalysis required for this structure, most proposals focus on the structural shift entailed in detaching the initially ambiguous noun phrase from its first thematic assigner (the subordinate verb "hunted" in (3a,b) and reattached it to the matrix verb (in this case, "hurried"). For Fodor and Inoue, this detachment is achieved by an operation dubbed "theft", which may be preceded by a "tug of war" between the two verbs (Fodor & Inoue, 1998, p. 122). Others conceptualise the restructuring process in terms of competition either associated with conflicts resulting from overlapping or interacting thematic processing domains (Ferreira & Henderson, 1998) or because of competition between incompatible pairs of attachment nodes (Stevenson, 1998). In each case, three linguistic entities are assumed to play a central role in the process of repair: (i) the disambiguating verb itself (on the grounds that this becomes the final attachment site for the noun phrase once the repair has been completed); (ii) the subordinate verb—its early competitor for the noun phrase argument and eventually the *detachment* site and (iii) the head of the disputed noun phrase itself. Assuming that the eyes are dispatched to fixate on objects that are the focus of current attention (as convincingly demonstrated by work using the Visual World paradigm—e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy,1995), the most plausible prediction from each of these accounts is that the eyes should be sent back either to the detachment site (i.e., the subordinate verb: "hunted" for 3a,b) or to the head of the disputed noun phrase: "moose" (or, just conceivably, to any string of words between the two). (Of course, fixations on the disambiguating verb might equally well be predicted, but since such fixations would not count as regressions they offer no basis for making predictions about the distribution of regression landing sites.)

Thus, under Selective Reanalysis the best prediction we can make for unpunctuated sentences is that in order to repair the initial misanalysis, the eyes should be sent back to landing sites within words 4 or 6 for sentences like (3a) and to locations 9 or 11 for (3b). In contrast, in the punctuated control sentences neither peak should be as prominent (on the grounds that the insertion of the comma in these conditions should largely prevent initial misanalysis, obviating the later need for reanalysis and repair).

### Method

#### Participants

32 volunteers participated. All were students at the University of Exeter and aged between 18 and 23. Nine were male and 23 were female. None had participated in Experiment 1. Payment or course credit options were exactly the same as in Experiment 1. In this case 21 of the participants were paid in cash.

#### Materials

The materials consisted of the late-linebreak versions of the Early Misanalysis Area sentences used in Experiment 1 together with a new set of 24 Late Misanalysis Area sentences listed in the Appendix. Each "Early" sentence was converted into a "Late" sentence by removing the 5-word relative clause ("that was adj1 and adj2" and inserting before the pre-posed clause a 5-word expression like: "The adj noun verbed that.." The choice of individual replacement words was designed to preserve a close match across the Early- and Late-versions of all sentences for each successive word appearing in Line 1. As in Experiment 1, corresponding words in each position were matched within two (or very occasionally three) characters across the Early and Late Misanalysis Area conditions, and cumulative mismatches were avoided to ensure that word-breaks appeared in equivalent positions across the two sentence-forms. More precisely, working from an index position at the end of Word 11, if the new Word 11 (which was identical to the old Word 6) was at the upper end of the 5-7 character range, efforts were made to ensure that Word 9 was at the lower end of its own 5-8 letter range, thereby shifting the Word8/Word9 gap closer to its median position (16 characters to the left of the index point). Needless to say, this close match in layout across all conditions depends upon a formulaic use of word lengths and the full set of materials had to be constructed prior to running Experiment 1. Each base sentence represented a particular temporary ambiguity resolved at Word 12. For example in Sentence (3a,b) above the ambiguous phrase "the moose" is first linked to "hunted" and later relinquished to "hurried". Across the experiment as a whole, each base form (e.g., "hunted/moose/hurried") was presented in the Early and Late versions half the time. In 50% of each set the sentence was presented with a comma just before the ambiguous noun phrase (as illustrated in 3a,b and in the Appendix).

As in Experiment 1, there were 24 object-resolved two-line foils and 80 more varied fillers, again with commas used in half the sentences.

All other aspects of the study were identical to those for Experiment 1. The font was the same as were all aspects of the procedure.

### Results

The overall error rate for the comprehension questions was 9.3%. As for Experiment 1, we carried out individual analyses for the first twelve words as well as for Region 13 (comprising words 13 and 14 combined). In each case, the data were entered into two-way ANOVAs with repeated measures for both Comma (present or absent) and Type of Sentence (featuring Early potential misanalysis area: Words 4–6 versus Late misanalysis area: Words 9–11). As before, for Word 1–11 only a broad summary is reported here. Like Experiment 1, analyses across these early parts of the sentence showed evidence of increased processing load when the comma first made its appearance. This showed up as a marginal increase in regression rates at Word 4 for the Early misanalysis area condition and a corresponding and fully reliable increase at Word 9 in the Late misanalysis area condition. Beyond this, there were some isolated effects linked to the fact that lexical content differed for Early and Late sentences. For example, at the very first word of the sentence, first-pass time was reliably shorter in the Late condition (where this word was almost always "the") than in the Early condition (where it was a conjunction such as "while"). In relation to total reading time, there were effects at all words other than Words 4, 6, 7 and 11. However, since total reading time effects in the absence of initial effects largely reflect later reinspections of the words in question, we pursue our analysis of these phenomena below by examining the details of regressions launched from the disambiguation word.

We turn next to examine the eye-tracking behavior at Word 12 and the following two words (Region 13), with a special emphasis on data patterns likely to be associated with disambiguation effects on the vicinity or Word 12. Seven of the 32 participants skipped Word 12 (and six skipped Region 13) for all six test sentences during at least one of the four sub-conditions, and all data for these participants were therefore discarded for their respective analyses. Table 2 provides a summary of the means, F-ratios, and confidence intervals for each of the standard eye-tracking measures in each of these locations. As expected, these data showed a range of main effects of punctuation, signalling increased processing difficulty in the No Comma condition for which disambiguation effects were anticipated. At Word 12 these effects emerged for all five measures for the materials analyses, but in analyses by participants only the first pass, regression path duration and total reading time effects were significant. There was a hint that these effects spilt over into the regression-based measures in Region 13, but as these delayed effects were not reliable we focus our attention on processing at Word 12. On the disambiguation word itself, there were no reliable main effects of sentence Type and no Comma

**Table 2**
Mean reading times and regression rates across the Comma and No Comma versions of sentences with Early and Late Misanalysis Areas

| | Mean eye-tracking measure for: No Comma/Early area; No Comma/Late area; Comma/Early area; Comma/Late area | | *F*-ratio for (in order): Punctuation Misanalysis area; Punctuation × Misanalysis area | | | 95% Confidence interval |
|---|---|---|---|---|---|---|
| *Word 12* | | | | | | |
| First Fixation (in ms.) | 229 | **P** | $F_1(1,24) = 2.74$ | $F_2(1,23) = 4.85^{*}$ | | 26 |
| | 228 | **M** | $F_1(1,24) < 1$ | $F_2(1,23) = 1.85$ | | 20 |
| | 210 | **P×M** | $F_1(1,24) < 1$ | $F_2(1,23) = 3.22$ | | 21 |
| | 205 | | | | | |
| First Pass time (in ms.) | 308 | **P** | $F_1(1,24) = 6.20^{*}$ | $F_2(1,23) = 7.42^{*}$ | $\mathrm{Min}F'(1,46) = 3.48$ | 30 |
| | 278 | **M** | $F_1(1,24) = 1.40$ | $F_2(1,23) < 1$ | | 27 |
| | 257 | **P×M** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 34 |
| | 257 | | | | | |
| Regression path duration (in ms.) | 485 | **P** | $F_1(1,24) = 6.45^{*}$ | $F_2(1,23) = 12.58^{**}$ | $\mathrm{Min}F'(1,43) = 4.26^{*}$ | 99 |
| | 482 | **M** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 108 |
| | 367 | **P×M** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 75 |
| | 356 | | | | | |
| Total reading time (in ms.) | 314 | **P** | $F_1(1,31) = 10.55^{**}$ | $F_2(1,23) = 26.11^{**}$ | $\mathrm{Min}F'(1,50) = 7.51^{**}$ | 50 |
| | 306 | **M** | $F_1(1,31) < 1$ | $F_2(1,23) = 10.93$ | | 49 |
| | 244 | **P×M** | $F_1(1,31) < 1$ | $F_2(1,23) < 1$ | | 32 |
| | 216 | | | | | |
| Regression rate (%) | 21.9% | **P** | $F_1(1,24) < 1$ | $F_2(1,23) = 4.50^{*}$ | | 13.1% |
| | 27.7% | **M** | $F_1(1,24) < 1$ | $F_2(1,23) < 1$ | | 8.6% |
| | 23.5% | **P×M** | $F_1(1,24) = 2.93$ | $F_2(1,23) = 2.11$ | | 6.7% |
| | 18.2% | | | | | |
| *Region 13* | | | | | | |
| First Fixation (in ms.) | 187 | **P** | $F_1(1,25) = 1.43$ | $F_2(1,21) = 2.67$ | | 21 |
| | 203 | **M** | $F_1(1,25) < 1$ | $F_2(1,21) < 1$ | | 23 |
| | 188 | **P×L** | $F_1(1,25) = 1.33$ | $F_2(1,21) < 1$ | | 24 |
| | 177 | | | | | |
| First Pass time (in ms.) | 325 | **P** | $F_1(1,25) = 2.68$ | $F_2(1,21) = 5.65^{*}$ | | 31 |
| | 332 | **M** | $F_1(1,25) < 1$ | $F_2(1,21) < 1$ | | 26 |
| | 300 | **P×M** | $F_2(1,21) < 1 F_1(1,25) < 1$ | 25 | | |
| | 308 | | | | | |
| Regression path duration (in ms.) | 625 | **P** | $F_1(1,25) = 10.35^{**}$ | $F_2(1,21) = 3.79$ | | 101 |
| | 389 | **M** | $F_1(1,25) = 8.51^{**}$ | $F_2(1,21) = 5.48^{*}$ | $\mathrm{Min}F'(1,42) = 3.33$ | 98 |
| | 370 | **P×M** | $F_1(1,25) = 4.07$ | $F_2(1,21) = 1.28$ | | 100 |
| | 329 | | | | | |
| Total reading time (in ms.) | 355 | **P** | $F_1(1,31) = 3.45$ | $F_2(1,23) = 3.94$ | | 43 |
| | 348 | **M** | $F_1(1,31) < 1$ | $F_2(1,23) < 1$ | | 34 |
| | 306 | **P×M** | $F_1(1,31) < 1$ | $F_2(1,23) < 1$ | | 24 |
| | 320 | | | | | |
| Regression rate (%) | 18.7% | **P** | $F_1(1,25) = 8.11^{*}$ | $F_2(1,21) = 3.79$ | | 6.5% |
| | 6.8% | **M** | $F_1(1,25) = 4.81^{*}$ | $F_2(1,21) = 3.01$ | | 6.9% |
| | 5.1% | **P×M** | $F_1(1,25) = 1.99$ | $F_2(1,21) < 1$ | | 6.6% |
| | 2.4% | | | | | |

Also listed are the *F*-ratios and 95% confidence intervals for the two main effects and their interaction. The top half of the table gives the results for Word 12 (the disambiguating word) and the remainder covers Region 13 (the post-disambiguation region). A single asterisk indicates $p < .05$, and a pair indicates $p < .01$.

by Type interactions. This indicates that at this point, at least for the measures highlighted here, there were no marked processing differences as a function of shifts in the point of potential misanalysis earlier in the sentence. Immediately beyond this, in Region 13, there was a regression path duration Type effect indicating that the regression sequences launched at this point were more protracted when the misanalysis area was in the Early location. Numerically this effect was more pronounced for unpunctuated materials, but as the Comma by Type interaction was not reliable, the results do not provide solid evidence that the increase is concentrated in this condition. Overall, these analyses reveal clear indications that readers encounter increased processing difficulties when they reach the disambiguating word but there is little evidence that the form of processing is altered according to whether the initial misanalysis occurred just a word or two back or much nearer to the beginning of the sentence.

To probe in more detail for effects of this kind we now proceed to examine the first (and later) landing sites of regressions launched from Word 12. In the first the analysis we employ the approach described earlier to generate regression signatures for saccades launched from Word 12 for the 2 (Comma versus No comma) by 2 (Sentence Type: Early vs. Late misanalysis region) sub-conditions. These signatures are shown in Fig. 3. As in the case of the equivalent analysis in Experiment 1, there was substantial data loss, with 22 of the 32 participants and 18 sentences being excluded from the analysis because they did not contribute data to all four sub-conditions. As before, the ANOVA showed a very pronounced main effect of Landing Site ($F_1(10,90) = 36.55$, $p < .001$; $F_2(10,50) = 19.39$, $p < .001$; Min$F'(10,101) = 12.67$, $p < .001$). However, there was no Landing site by Punctuation interaction ($F_1 < 1$; $F_2(10,50) = 2.10$, $p < .05$). Crucially for evaluation of the Selective Reanalysis hypothesis, there were hints of regression signature changes in the form of a marginally reliable triple interaction ($F_1(10,90) = 2.87$, $p < .05$; $F_2(10,50) = 1.74$, $.05 < p < .1$). To test for evidence of differential returns to the Early misanalysis area (i.e., Words 4–6) with one set of sentences and the Late area (Words 9–11) in the other, we compared summed returns to Word 4, 5 and 6 with the total of all regressions to Words 9–11. Within the overall triple interaction, this subcomponent of the Landing Site factor showed at least a *suggestion* that regressions were directed back to the expected sentence regions. In the Early misanalysis area condition, using as a measure returns in the No Comma condition minus returns in the Comma condition, there were 30.6% more immediate excess visits to Words 4-6 in the No comma condition than there were to Words 9-11, whereas in the Late misanalysis area condition the pattern was reversed (32.4% *fewer* excess visits to the Early than to the Late area) ($F_1(1,9) = 7.05$, $p < .05$; $F_2(1,5) = 3.96$, $.1 < p < .15$).

This failure to provide a robust demonstration of selective returns may be due to the exclusive attention paid to the landing site of the very first regressive fixation. Across all participants for the two unpunctuated conditions, the average length of regressive sequences was a little more than 5 fixations, and it is worth asking whether the full sets show evidence of selective visiting patterns. To explore this possibility, we tabulated the landing sites for each participant (or sentence) and for *all* fixations falling within the regression sequence (i.e., the first regression plus all other fixations prior to any movement to material beyond Word 12). Fig. 4 plots the distribution of the total number of such regressive visits to the eleven different word positions across the 2 × 2 sub-conditions of the experiment. The overall pattern is broadly similar to that seen for first regressions in Fig. 3, with a heavy preponderance of visits to words close to the launch site. However, in this case there appears to be more marked evidence of differential returns to the different misanalysis areas. To examine the pattern more closely, Fig. 5 provides a plot for each word of the *additional* number of regressive visits for the unpunctuated compared with the punctuated control condition. This appears to how that syntactically triggered

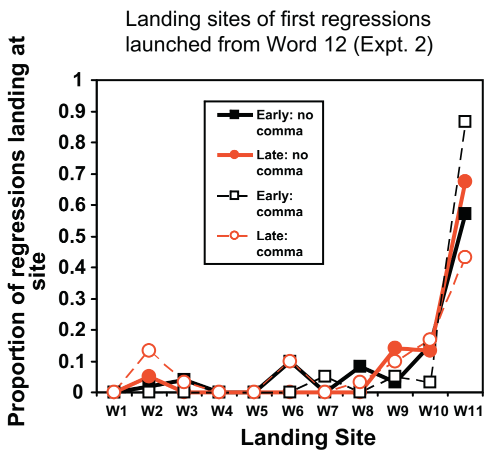


**Fig. 3.** Experiment 2 Regression Signatures for first regressions from Word 12 in each of the four sub-conditions (Comma versus No comma crossed with materials with Early and Late areas of potential misanalysis).

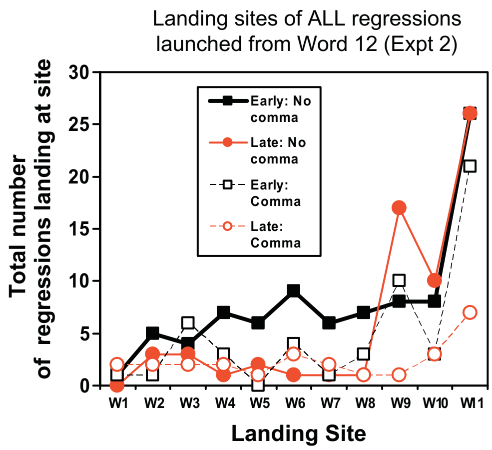


**Fig. 4.** A tally of all landing sites visited during regression sequences launched from Word 12. Counts are plotted separately for each of the 2 × 2 experimental conditions.

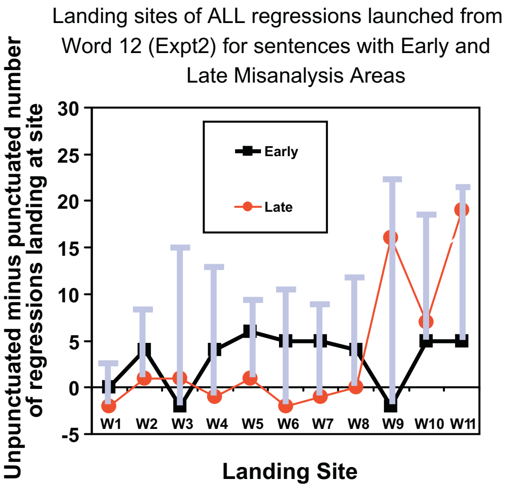


**Fig. 5.** Plots of the additional visits to different landing sites in the No comma condition (assumed to reflect the properties of syntactically triggered regressions). The *y*-value is the total number of regressive visits in the No comma condition minus the corresponding tally for the Comma condition. There are separate graphs for the two types of sentence. For sentences with a region of potential misanalysis in the Early position, Selective Reanalysis predicts a concentration of returns to Words 4 and 6. For those with a Late misanalysis region regressions are expected to gravitate toward Words 9 and 11.

regressions tend to concentrate round Words 4–6 for the Early Misanalysis Region materials and to gravitate toward Word 9–11 for sentences with Late Misanalysis Regions. As the error bars indicate, no individual word attracted statistically different numbers of punctuation corrected revisits across the two sentence types. However these individual confidence intervals fail to capture the fact that the return patterns flip (in the predicted manner) between the Early and Late Misanalysis regions. To provide a formal test of these essentially interactive effects, we calculated by participants (and also by sentences) the total number of return visits to each of the first eleven words during regression sequences initiated from Word 12. These 11-argument vectors were subclassified over the $2 \times 2$ experimental conditions and submitted to 3-way repeated measures ANOVAs with factors comprising Landing Site (11 levels, corresponding to Words 1–11), Comma (2 levels) and Sentence Type (Early vs. Late misanalysis area). This yielded a Comma main effect showing that in total there were more regressive fixations in the No Comma condition ($F_1(1,31) = 7.61$, $p < .01$; $F_2(1,23) = 7.29$, $p < .05$; Min$F'(1,52) = 3.72$, $.05 < p < .1$). There was no main effect of Sentence Type ($F_1,F_2 < 1$), but there was a very strong effect of Landing site ($F_1(10,310) = 23.59$, $p < .001$; $F_2(10,230) = 31.13$, $p < .001$; Min$F'(10,539) = 13.42$, $p < .001$). Sentence type did not interact with Comma or Landing site (($F_1,F_2 < 1.4$, in both cases). However, there was a 2-way interaction between landing Site and Comma ($F_1(10,310) = 3.58$, $p < .001$; $F_2(10,230) = 3.17$, $p < .001$; Min$F'(10,516) = 1.68$, $.05 < p < .1$). Finally, and most critically for inferences concerning selectively of regressive returns, the triple interaction was reliable ($F_1(10,310) = 4.84$, $p < .001$; $F_2(10,230) = 5.34$, $p < .001$; Min$F'(10,534) = 2.54$, $p = .005$).

This triple interaction shows that, as depicted in Fig. 5, that the overall pattern of differential returns to potential regressive landing sites alters according to the type of sentence on display. Informally, and entirely in line with the predictions of Selective Reanalysis there seems to be a preponderance of visits to Words 9, 10 and 11 in the Late Misanalysis Area condition. That is, when they make regressions from Word 12, readers seem to concentrate their additional fixations (relative to the control) within the region of potential misanalysis. In the Early Misanalysis Area condition the distribution of surplus regressive visits was not quite to clear-cut. There was some indication of increased attendance at most early and middle landing sites. As set out in the introduction to the current experiment, Selective Reanalysis predicts a particular concentration of returns to Words 4, 5 and 6. To test this specific prediction, we extracted a component of the triple interaction by running a contrast that restricted the Landing Site factor to a comparison between the total number of returns to Words 4, 5 and 6 on the one hand and, on the other, the total number of revisits to Words 9-11. On the null hypothesis that the two Sentence Types trigger an equivalent balance of returns to these two areas this particular subcomponent of the triple interaction would not be expected to reach significance. In fact, the contrast yielded a very solid effect ($F_1(1,31) = 11.73$, $p < .01$; $F_2(1,23) = 26.18$, $p < .001$; Min-

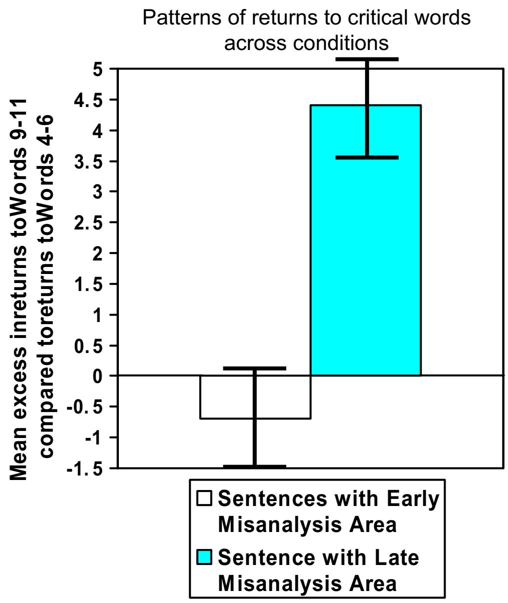


**Fig. 6.** The data used in a formal test of Selective Reanalysis. For each participant we calculated the following value: (Number of returns to Words 9–11 in the No comma condition *minus* number of returns to Words 9–11 in the Comma condition)–(Number of returns to Words 4-6 in the No comma condition *minus* number of returns to Words 4–6 in the Comma condition). The left-hand (negative valued) bar gives the mean of this score for sentences with an Early potential misanalysis region, and the right-hand bar gives the corresponding value for sentences with Late misanalysis regions. The score is taken to represent the degree to which Words 9-11 attract *more* syntactically triggered regressive visits than is the case for Words 4 to 6.

$F'(1,51) = 8.10$, $p < .01$). Fig. 6 provides a plot showing how the use of the different Sentence Types alters the differential tendencies to revisit these areas.

## Regression-contingent reading time effects

In a final set of analyses intended to revisit the Altmann et al. (1992) findings concerning regression-contingent reading times, we combined data from both experiments to establish whether the present experiments corroborate the observation that reading times preceding regressions are shorter than those terminated by a progressive saccade. For each participant (sentence) in each experiment we recorded the mean first-pass reading time pooled across Words 2–12 (a) for first-pass episodes ending in a regression and (b) for visits culminating in a progressive movement. (The data for Word 1 were excluded because regressions from this position were by definition impossible.) The results were entered into a 2-way mixed ANOVA with a non-repeated factor for Experiment (Experiment 1 or Experiment 2) and an Exit Type (regression versus progression) being repeated for the materials analysis and non-repeated for the participant analysis. The results provided very solid confirmation of the earlier findings. Overall, the Exit Type main effect showed that the first-pass reading times were 26 ms. faster when they preceded a regression (232 ms.) than when the next saccade was a progression (258 ms.) (CI ± 9 ms.) ($F_1(1,58) = 19.64$, $p < .001$; $F_2(1,23) = 35.50$, $p < .001$; Min$F'(1,78) = 12.65$, $p < .001$). There was no sign of an Experiment main effect ($F_1 < 1$; $F_2 < 1.6$), nor any interaction between the two factors ($F_1 < 1$; $F_2 < 1.8$). Finally, in a further refinement to this analysis we narrowed our focus to trials in which there was a single fixation on a word before either a progression or regression. (Trials in which there were two or more successive fixations on the word were treated as missing data for the purposes of this analysis.) Here again, the same regression-contingent reading time effect re-emerged, albeit in a slightly weakened form. In this case, fixations followed immediately by a regression (mean: 204 ms.) remained reliably faster than those preceding a progression (mean: 220 ms.: CI ± 8 ms) ($F_1(1,58) = 9.48$, $p < .01$; $F_2(1,23) = 44.70$, $p < .001$; Min$F'(1,76) = 7.82$, $p < .01$). As before, there was no sign of an Experiment main effect ($F_1 < 1$; $F_2 < 2.2$), nor any interaction between the two factors ($F_1 < 1$; $F_2 < 1.8$).

### *Discussion*

After detailed analysis, the results of the experiment showed clear evidence that regressive eye-movements can be placed under some kind of linguistic control. However, the pattern of results was somewhat at odds with the precise predictions of the Selective Reanalysis hypothesis. As previously indicated, in its original form the hypothesis stated that the eyes are sent "directly to [..] the region containing the information that would permit the parser to locate the source of its error" (Frazier & Rayner, 1982; p.188). Against this, we were unable to find solid statistical support for the claim that the very first regression was sent to what we have taken to be the critical part of the text. Granted, there were trends in this direction and arguably the effect could well reach conventional levels of significance if we were to repeat the experiment with increased numbers of sentences to boost the power of the materials effect. Notwithstanding our failure to find good evidence for *direct* returns, by tracking the itineraries of the remaining fixations within each regression sequence we were able to demonstrate that, in time, fixations do tend to concentrate in the expected areas of the sentence. Informally, what seems to happen is that when the target area is a little remote, instead of locating it instantly the eye-control system seems to program saccades to land on one or two closer words and use them as 'stepping stones' *en route* to the eventual destination. More concretely, simple tallies show that with the Early Misanalysis Area materials, only 30% of "successful" revisits to Words 4 or 6 were achieved after the very first saccade. Half of all eventual hits required at *least* three fixations to reach the target word (with the maximum lag in this particular data set being 13 fixations before the target was finally located). So, whereas the overall tallies indicate unequivocally that there is selective guidance to particular areas of the sentence, this guidance seems to take the form of a series of successive steps rather than a "direct" jump to the target area. In this sense, they bear a resemblance to the patterns of regressions reported in a question answering task reported by Inhoff and Weger (2005). Attesting further to the relative imprecision of the system, it is also worth noting that for Early Misanalysis Area materials only 22.7% of all regression sequences included at least one fixation on one of the target words. As a comparison, in 41.9% of cases the regressions went back no further than Word 11 (and in a further 23.5% of cases the closest approach was to Word 9).

In short, although the results of the present experiment provide unequivocal support for the notion that regression scan-paths are subject to linguistic influences, the picture that emerges is not one of assertive control as envisaged by Selective Reanalysis. Instead, the detailed examination of our regression sequences suggest a tentative and somewhat inefficient form of control: one in which the target area is actually never even *reached* in the majority of regression scan-paths. Taken together with demonstrations of format effects in Experiment 1, this encourages the view that the linguistic processor is a relatively low status driver for regressions and that its instructions are frequently overridden or swamped by those issued by other competing systems.

In a side analysis, the results provided very clear-cut corroboration for the previously reported regression-contingent reading time effects (Altmann et al., 1992; Rayner & Sereno, 1994). This effect occurred in the face of a recently reported finding that for at least a subset of regressions the bias may be in the opposite direction. Rayner et al. (2003) found that fixations are lengthened when they precede returns to words previously fixated one step earlier in the sequence. Our regressions are known to have included sequences of this kind. But the overall regression speeding suggests that any inhibition on trials of this kind must have been swamped and reversed by samples of rapidly triggered regressions on other trials. Both first-pass reading times and single fixation times were shorter when they immediately preceded regressions. Although there is no reason to suspect that this effect is specifically linked to linguistic guidance, making it less central to our arguments than the scan-path data discussed elsewhere, it remains an intriguing finding that has yet to receive proper theoretical attention. It certainly seems to be the case that a comprehensive account of the linguistic control of eye-movements would need to explain why regressions appear to be launched at an earlier point in processing than is the case for progressions, an observation that is in apparent conflict with the overwhelming bulk of evidence that regressive movements are typically deployed when there is an *increased* amount of work to be done. Some of the implications of this finding will be explored in the General Discussion below.

## General discussion

The main purpose of the present study has been to examine how eye-movements, and most particularly regressive eye-movements, are influenced by syntactic operations such as those associated with the resolution of the temporal structural ambiguity. The primary reason for dwelling on this is that eventual success in mapping the details of the eye-control system is likely to be a prerequisite for developing accurate computational models of syntactic processing (as calibrated by eye-tracking data). In this sense, the present study can be seen as part of the groundwork for a broader exercise of modelling syntactic processing.

By recording eye-movements as people read specially crafted sentences, we compared two competing visions of regression-control. The first assumes that regressions are tightly articulated with linguistic operations (as exemplified by Frazier & Rayner, 1982, Selective Reanalysis hypothesis). The second is that they are merely used to buy time for the linguistic processor to catch up with its work (the Time Out hypothesis). In the event, the results showed that neither of these (admittedly caricatured) positions turned out to be fully compatible with the data. Against exclusively linguistic control, a range of layout effects in Experiment 1 showed that syntactically driven regressions are subject to other extraneous influences. Also questioning this position is the evidence from Experiment 2 that regression guidance is characterized not by immediate and precise deployment but by tentative, step-by-step approaches to the target area. Equally, however, the Time Out hypothesis can be ruled out comprehensively as a complete account of regressive eye control. The outcome of Experiment 2 would have been entirely different if it had been the case that the function of regressions had simply been to prevent *pro*gression and so to buy time for an overstretched linguistic processor to complete its work. Notwithstanding the apparent imprecision of the control, the results of this experiment show definitively that at least a proportion of regression scan-paths are guided to areas of linguistic significance. Systematic patterns of this kind would not occur with a system exclusively concerned with holding back the flow of new work to the linguistic processor. Taken together, the most obvious interpretation of these findings is that, in contrast with the patterns that emerge in other tasks (e.g., Kennedy & Murray, 1987; Murray et al., 1988), regression sequences triggered by syntactic disambiguation are placed under relatively *loosely-coupled* control with the guidance of successive fixations being shared between both linguistic and non-linguistic mechanisms.

In certain important respects these conclusions are at odds with claims made on the basis of earlier studies examining these issues. As far as we are aware, previous investigators have never had to resort to tracking the full trajectories of regression sequences in order to secure good evidence for selective returns. In the past, inferences have been drawn exclusively from data concerning the destination of the very first regressive fixation launched from the disambiguation area (Frazier & Rayner, 1982; Meseguer et al., 2002), This raises the question of why the regressive sequences generated by the participants in our experiments appear to have been so much less tightly controlled than the patterns obtained in previous work.

In the case of comparisons with the findings reported by Frazier and Rayner (1982), what emerges is that the discrepancies are nothing like as stark as they might seem. In tabulating regressions launched out of the disambiguation region, Frazier and Rayner (1982) distinguished between three broad landing areas, labelled "Beginning of sentence", "Before ambiguity" and "In ambiguity", with this last area being directly adjacent to the launch site. The results showed that numerically (though not statistically) the largest number of the regressions were directed into this (i.e., the closest) area. However, what this study did not do was provide statistical evidence that this tendency was accentuated when the launch area included an unexpected continuation compared to controls in which it contained more predictable material. For any individual regression, the most likely landing site is the word immediately to the left of the launch point. Indeed, on the evidence of the present experiments this may be particularly true of non-syntactically triggered regressions. By using a misanalysis area which happened to be immediately to the left of the launch-site, Frazier and Rayner (1982) created a situation in which it was difficult to disentangle the effects of different kinds of regression. Put simply, it is impossible to tell from the data whether the regressions were sent to the "In ambiguity" area because they were responding to instructions from a syntactic driver or merely because they were going back to the preceding word. Furthermore, the very proximity of the misanalysis area actually provides a very weak test of direct targeting. In a linguistically guided system it is quite possible that a regression can be sent "directly" to a target that happens to be a word or two away, but that two or more steps are required in cases where the target word is more remote. In short, and admittedly with the benefit of substantial hindsight, there is little that can be pointed to in Frazier and Rayner (1982) data that firmly supports the notion of tightly coupled control.

We turn next to comparisons with the more recent study by Meseguer et al. (2002). In this case the potential sources of the discrepancies between the two studies is less obvious. It will be recalled that Meseguer et al.

(2002) reported evidence of precise, word-specific regressive returns for the very first saccade launched back from the disambiguation region. In their study, the target area was located at a distance of several words from the launch site and, indeed, on a separate line in some conditions. In these respects, there are no grounds for expecting to see greater precision in their experiment. On other counts the two studies are difficult to compare. Meseguer et al. (2002) used different structural ambiguities (in a different language), different calculations to quantify reinspection rates, and—as indicated earlier—they also compounded returns from the disambiguation region with any that might have been launched from the end of the sentence. It may be (as has been suggested by Chuck Clifton, personal communication) that the linguistic structures used in the earlier study afforded more precise information about the physical location of earlier analysis. It is also possible that the Meseguer et al. (2002) regressions were somewhat more erratic than suggested by the published account of the data. Unlike the tack taken in the present study, these authors did not track the later trajectories of regression sequences. It is possible that amongst the many regression sequences that did *not* start directly with a visit to the misanalysis area there were substantial numbers that took two or three fixations before they gravitated toward this target area. In other words, had the authors extracted comparable measures to those used in the present study, it is conceivable that they might have used different language to describe their findings. A more extensive examination of regression phenomena will be required before researchers are in a position to make definitive statements about the kind of precision that can be expected with different linguistic structures and under different experimental conditions.

To take stock, then, at this point we have evidence that the trajectories of regression sequences are subject to disparate forms of control. Experiment 1 showed that the choice of landing sites was influenced in a variety of ways by physical details such as a layout of the text. In Experiment 2, where physical appearance was controlled, it proved possible to detect forms of guidance that can only plausibly be linked to some aspect of linguistic processing. It is beyond the scope and ambition of the present paper to specify how these different forms of control might be co-ordinated and reconciled with one another. But even in the absence of a detailed specification of the mechanisms for regression control, these findings have direct implications for several existing theories of sentence processing, and we turn next to these more theoretical concerns.

## Reconciling the findings with models of eye-movement control

How well are the present data handled within current models of eye-movements in reading? The blunt answer seems to be that none of the existing models is well equipped to account for the full range of phenomena described here. In a brisk survey to tackle this question, it is useful to distinguish between models that are explicitly designed to explain parsing effects and those that are not. The first class includes models like the Competition Integration model (McRae et al., 1998) and its variants (e.g., Green & Mitchell, 2006; Spivey & Tanenhaus, 1998), other specific models using moment-to-moment competition (e.g., Stevenson, 1993; Stevenson, 1998), retrieval based models (e.g., Lewis & Vasishth, 2005), Bayesian models (e.g., Narayanan & Jurafsky, 2002), as well as a wider class of models based on the use and training of Simple Recurrent Networks (e.g. Christiansen & Chater, 1999; Christiansen & Chater, 2001; Konieczny & Döring, 2003; MacDonald & Christiansen, 2002; Rohde, 2002 and Tabor et al., 1997). In each case the model can be used to generate a metric intended to reflect the segment-to-segment processing load in performing the structural analysis. On certain occasions, these metrics have been explicitly linked to specific eye-tracking observations (e.g., Konieczny & Döring, 2003; Spivey & Tanenhaus, 1998). In other cases, the connection has not been set out as such. However, given that the models uniformly seek to account for standard observations (e.g., disambiguation effects), it is perhaps not unreasonable to infer that the authors premised their arguments on the assumption that the outputs of their models must somehow determine the patterns of holds and progressions of the eyes. With the exception of the Lewis and Vasishth Activation/Retrieval model (which includes provision for timing and decay effects), the input for the load-predicting calculations is restricted either to information about the string of words comprising the test sentence or to a sequence of linguistically-based probabilities or biases. None of the models offers any way of specifying that a given input word (or bias) is associated with material at the beginning or end of a line, or any other comparable spatial information. It follows that none is capable of accounting for any of the Line-break effects demonstrated in Experiment 1. Conceivably some could be construed not as accounts of parsing processes in reading, but as models of syntactic analysis for auditory input. In that case the failure to handle format effects may not be seen as a serious shortcoming of the model. In other cases there may be a tacit assumption that the goal of the work is to model reading measures only (say) for material that appears in the middle of a line of text. It is not clear whether high priority should be given to adapting any of these models to handle the ways in which observed measures are changed by format effects. However, what is clear is that if such measures, left uncontrolled, are subject to dramatic variation (as demonstrated in Experiment 1), then there can be little confidence that any data-fitting exercise will achieve anything more than parameterizing the model to fit different degrees of format-generated noise.

If existing high-level models are poorly equipped to account for format effects, most are no better placed to deal with linguistically guided regressions, and there have been no prior attempts to model these effects computationally. Indeed, most accounts seem distinctly ill-prepared to handle such phenomena. To the extent that these models have been designed to produce a numerical fit to human data, the focus of attention has almost invariably been the challenge of providing fits for the word-by-word or phrase-by-phrase fluctuations in reading time. Most models do this by generating a measure of processing difficulty either based on the number of cycles or iterations to complete a settling process (e.g., McRae et al., 1998; Spivey & Tanenhaus, 1998; Stevenson, 1993; Stevenson, 1998; Tabor et al., 1997), or a discrepancy (error) calculation (e.g., Christiansen & Chater, 1999; Christiansen & Chater, 2001; MacDonald & Christiansen, 2002) or some other explicitly constructed metric intended to reflect processing difficulty (e.g., Konieczny & Döring, 2003; Levy, 2008; Lewis, 1993; Lewis & Vasishth, 2005; Narayanan & Jurafsky, 2002; Rohde, 2002). In no case does the metric offer any insight as to where the eyes should be directed as the operation runs to completion. For example, in the Competition Integration model (McRae et al., 1998; Spivey & Tanenhaus, 1998; Tanenhaus et al. 2000), the time interval between first encountering a phrase and moving on to later material is simulated by the number of cycles it takes a settling process to adjust an activation unit to a threshold level. Applied to eye-movement data it has nothing to say about where the eyes might be directed during the course of this settling process. Indeed, lacking a mechanism for storing the location of past material, it excludes even a representation that could potentially be used to specify the co-ordinates of any such fixations. Arguably, such shortfalls could be rectified by positing such a representation and adding also a pointer to direct the eyes from moment to moment. But ultimately, for such a system to generate different scan-paths in different situations (as in Experiment 2) it would have to be acknowledged that there is a qualitative difference between one settling process and another and, given existing mechanisms, it is difficult to see scope for finding any plausible way of generating such differences. Similar difficulties emerge for other implemented models. They simply have no account to offer for the fact that regression sequences follow different trajectories with different materials.

One move that might be adopted in certain cases would be to treat regressive eye-movements as being epiphenomenal. For example, modelers might follow Weger and Inhoff (2007) "verbal reconstruction" hypothesis and assume that regressive fixations merely shadow mental operations already in train. In the case of models that do more than quantify the degree of predictive error and, for example, incorporate a settling process that tracks over time (or cycles) the relative status of competing representations, it is conceivable that a spin-off system could direct the eyes to the locality of a competitor that is gaining in prominence. So, in Experiment 2 if after disambiguation a representation for "NP-as-matrix-subject" gains in strength, then a spin-off system might send the eyes to the sentence area incorporating crucial entities (such as the NP itself). Of course, epiphenomenal accounts of this kind remain highly speculative. Moreover, it is worth pointing out that there are aspects of the present data that do not resonate at all well with the proposal that regressive fixations are a mere spinoff of an essentially verbal retrieval process. In Experiment 1, the pixel co-ordinates and all other dimensions of each pre-disambiguation word were identical across the two line-break conditions. Had participants set out to retrieve any particular word, with their eyes therefore being dragged to the visual co-ordinates of the stored representation, the regressions signatures would necessarily have been identical in the two cases. The fact that they were not indicates that spin-off accounts (like the verbal reconstruction hypothesis) are incapable of providing a complete description of the data. At the very least, the present experimental results indicate that if aspects of regression control are ceded to epiphenomenal operations, then these processes still need a great deal of further refinement before they come anywhere close to accounting for the data.

In weighing up the prospects for different theoretical accounts for 'regression scan-path modulation', it is also important to consider the classic proposal that regressions mark the instigation of non-standard syntactic operations: processes that are concerned not with the choice of a structural analysis, but with the dismantling and repair of analyses that have proved to be non-viable. It was precisely on the basis of such 'reanalysis' theorizing that our predictions were derived for Experiment 2 and, in light of their broad success in anticipating the results there is a strong case for exploring this line of argument. In practice, however, it proves very difficult to reconcile this kind of reconfiguration with the detailed machinery of current implemented models. Unlike some of the less formally expressed theories, existing implemented models are relentlessly forward-looking and none is furnished with the capacity to revert to an earlier configuration. This makes it impossible for them to test the potential of an altered processing itinerary. In particular one class of models may be singularly ill-equipped the handle backtracking discontinuities of this kind. The models in question are those that base their predictions directly or indirectly on the output of a previously trained connectionist network (e.g., Christiansen & Chater, 1999; Christiansen & Chater, 2001; Konieczny & Döring, 2003; MacDonald & Christiansen, 2002; Rohde, 2002 and Tabor et al., 1997). In all current models of this kind, training is based on exposing the network to legal input strings (with legality often defined in terms of a finite state grammar). This history of training ensures that the networks accumulate vast experience in dealing with words arriving in the "correct" (i.e., rule-governed) sequential order. However, they have no training at all preparing them to deal with words received in the *wrong* order. That is, they do not have the experience-shaped connection weights that would equip them to deal with order-violating backtracking moves. If a parser trained exclusively with strictly ordered inputs is presented with a regressive move to Word 6 (say) following Word 12, then it will have nothing in its repertoire preparing it to deal with this transition. In some cases the models may make numerical predictions but the outputs would be meaningless. Nor do any of the models have facilities for undoing the activation changes caused by the last few words, allowing the system to be reset to its configuration at the point of the regression target word. Because the retention of sequential discipline is so fundamental to models of this kind it is difficult to see how they could be adapted to handle backtracking operations.

We turn next to an examination of eye-movement and reading models that are not furnished with any kind of machinery for parsing sentences. To the degree that these models place emphasis on low-level spatial effects (as well as on lexical effects) they should, in principle, be well placed to handle format effects. However, some would have difficulty in dealing with the punctuation effects in our experiments. Note that the Word 12 and Region 13 materials showing Comma effects were separated by 7-8 words (or 40-50 characters) from the closest point on the left for which there was a physical difference between the alternative stimuli (i.e., the point at which the comma was inserted or excluded). Beyond Word 12 the two versions were completely identical. Models of "low-level" (i.e., lower than syntactic) eye-movement control tend to concentrate on predicting measures like fixation time as a function of the frequency of the fixated word and probabilities of single fixations, refixations or skipping of a word as a function of its length and frequency (e.g., Engbert et al., 2005; Reichle et al., 2003). To do this, they take as raw material the spatial array of letters comprising no more than four words or about 30 characters—often including the target word itself. Because of restrictions imposed by the use of relatively narrow attentional windows, punctuation differences 7–8 words back would never be represented as part of the current input in most of these models. In effect, the Comma and No-comma letter- and word-strings would be represented as identical inputs. It follows that for this aspect of the materials there is nothing to differentiate between the materials and therefore no basis for making differential predictions about eye-movement behaviour.

In accounting for the present data, the main deficiency encountered in most low-level models is that they fail to make any provision at all for eye-guidance by linguistic processes beyond the level of the word. However, in a recent extension of the E-Z Reader model (identified as E-Z Reader 10), Reichle et al. (submitted for publication) have added a new post-lexical "integration" process that can be characterized as serving this purpose. This is presented as a placeholder, covering a variety of different post-lexical processes. The operation is tasked with integrating each new fully-processed word with the prior text. In the current implementation there are two ways in which this activity is capable of exerting an influence on eye movements. Faced with very serious difficulties, the integration can simply fail, immediately triggering a saccade that overrides all other movements that may be in labile stages of preparation. In the present formulation, this saccade is sent to the disambiguating word with a parameterized probability, or to earlier material (with the complement of this probability). In the current realization, no statement is made about the internal spatial characteristics of the pre-disambiguation part of the distribution, but clearly any landing site from this set will classify the saccade as a regression. According to the model there is also a second way in which regressions can be triggered (this time as a result of "slow integration failure"). This occurs in cases where the integration process following Word N fails to reach its culmination prior to the completion of the two lexical processing operations associated with the following word (N + 1). On trials where integration lags in this way, the breakdown of smooth processing triggers, as before, a return probabilistically divided between the disambiguating word and earlier material (with the latter again producing regressions). While this model describes at least two different mechanisms by which post-lexical processing might exert an influence on eye-movements, it falls some way short of providing an adequate account of much of the current data on disambiguation effects. Most fundamentally, like earlier versions of the model, E-Z Reader 10 explicitly excludes from its explanatory domain eye movements made from words appearing at the beginnings and ends of lines. Strictly speaking, therefore, it does not claim to make predictions for any of the disambiguation word data described in the present study (because Word 12 was always in one of the two excluded line positions). Even if this restriction were lifted, there are several phenomena that would present varying degrees of difficulty for the model. First, it offers no obvious explanation for the fact that regression signatures change over different conditions (as summarized in Fig. 2, above). The model features just a single mechanism for responding to integration breakdown, and this provides no facility for explaining the fact that regressive saccades were projected greater distances in our unpunctuated than punctuated sentences. If both types of regression had been generated by the same mechanism, it is difficult to see why the regression signatures would end up being different for sentences presented with and without a comma. This lack of differentiation also leaves the model poorly prepared to deal with the systematic patterns in the words visited in the later phases of regression scan-paths (as shown in Figs. 5 and 6). In particular, the model offers no explanation at all for the fact that the regressions apparently gravitate to areas of the sentence in which the material was previously likely to have been misanalyzed. In addition to difficulties in accounting for what happens *after* regressions are launched, were a version of E-Z Reader 10 adapted to cover behavior at the ends of lines it would still have difficulty in explaining the format-linked changes in regression rates found in Experiment 1 above. Words at the ends of line are almost certain not to suffer from the second of the two forms of integration failure specified in the model (i.e., "slow integration failure"). This occurs when lexical processing of the following word reaches completion before the current word has been fully integrated. At the end of the line, the next word is positioned 60-70 characters to the left of the fixation point: well outside the range of any useful pre-saccadic lexical processing. The probability of parafoveal lexical processing reaching completion must therefore presumably be close to zero. As a consequence, according to the model, slow integration failure must be virtually non-existent in the end-of-line position. In contrast with this, the situation changes when the same word is shifted to the beginning of Line 2. Here, the following word could easily be processed parafoveally (almost always being a very short high-frequency word in our study). Presumably, this should markedly increase the probability of slow integration failure in this new position. Assuming that fast integration failure is not influenced by word position, the extended version of the model would seem to predict higher regression rates from a difficult-to-integrate word located at the beginning than at the end of a line—exactly the opposite of the solid finding obtained in our study. Some of these shortcomings of E-Z Reader 10 may be matters that can be addressed in future refinements of the model.

As an alternative to introducing high-level drivers, certain models (e.g., Engbert et al., 2005; Reichle et al., 1998) offer ways of handling supra-lexical effects by taking account of the predictability of upcoming material given all previous words in the sentence. In principle it is possible that a device of this kind could be used in an attempt to model our punctuation effects, as well as various other ambiguity resolution phenomena. What is less clear, however, is how such low-level models might account for the regression-contingent reading time finds of Altmann et al. (1992) (duly corroborated in the present study). The regression-contingent effects call for a mechanism in which fixation duration is in some way linked to events which take place after the fixation ends. The best prospects for explaining the data may lie in drawing upon accounts of time increases for increments that precede word-skipping saccades. Reichle et al. (2003) treat saccade preparation is a two-phase process, with the first stage being labile (and subject to cancellation when confronted with competing motor commands, as may happen when instructions for a word-skipping saccade emerge during the labile period). Such a mechanism can account for apparently retrospective effects: fixation latencies are predicted to be longer before a word-skipping saccade than for a more standard incremental advance (also see Engbert et al., 2005, p. 799, for an equivalent prediction). Note that in these systems latencies are increased when a "normal" or default progression from the current word (N) to N + 1 is overridden by a control-seizing event (here, the planning and execution of a word-skipping saccade). A natural assumption would be that the same data pattern would emerge in other default-overriding cases (e.g., replacing the N + 1 progression with a regression). However, this is the exact opposite of what we found. Our data showed that fixations followed by progressions (presumably the default) were actually reliably *slower* than those preceding the much rarer regressions. Within saccade-cancellation accounts, the only way of interpreting this finding seems to be to make the highly implausible assumption that the system treats *regressions* as the default and so incurs a penalty when some partially external system imposes an advancing saccade. In short, it seems unlikely that our regression-contingent findings can be interpreted as saccade-cancellation effects of the kind previously developed to account for word-skipping phenomena.

## Summary and conclusions

We started by observing that implemented models of sentence processing have not by and large treated it as part of their remit to account for the patterns of behaviour observed during regressive sequences of eye-movements. Traditionally these phenomena have been described within the framework of the "selective reanalysis hypothesis" (e.g., Fodor & Ferreira, 1998; Frazier & Rayner, 1982). According to this view, the eyes are seen to be under the tight control of operations deemed to be responsible for the revision and repair of discredited or non-viable structures. The constructs of "reanalysis" and "repair" are fundamentally incompatible with the mechanisms underpinning almost all currently implemented models of parsing. To the extent that this is the only viable explanation of the data, this entire class of quantitative models may therefore be seriously compromised. However, as noted in the Introduction, there are serious questions about the status of the existing evidence for selective reanalysis as an interpretation of regressive eye-movements. Very few studies have been conducted in the past, and those that have lack the manipulations required to rule out alternative explanations. One such account would be much easier to reconcile with the current generation of implemented models. This is the Time Out proposal (i.e., that regressions are programmed not to facilitate repair, but merely to buy time for the linguistic processor to catch up with its existing backlog of processing). If the data could be shown to be compatible with this account, then existing implemented models could be retained without the need for any radical re-engineering.

In the event, while our findings challenged many of the standard assumptions about selective reanalysis, the detailed scan sequences in Experiment 2 also revealed regression modulation effects that are impossible to reconcile with the Time Out account. Specifically, against standard assumptions, the results of Experiment 2 showed that on most trials regressions were not sent directly back to the critical area, typically taking 2-3 fixations to reach their intended target. In fact, detailed scrutiny of regression sequences showed that for remote target areas the majority of the scan-paths never reached the anticipated destination. It is possible that such sub-populations were under Time Out control. Also inconsistent with standard accounts was the finding in Experiment 1 that regressions were influenced by non-linguistic factors (i.e., those associated with the layout of the text). However, none of these qualifications should obscure the fact that in Experiment 2 we found also clear evidence that a linguistic manipulation altered the distribution of the landing sites of fixations within the regression sequence. Although lacking the predicted directness of control, these findings were entirely compatible with the true spirit of the Selective Reanalysis hypothesis. Crucially, they also raise serious problems for models that lack the machinery for sending the eyes to different parts of the sentence if and when sentence structures are changed. A challenge for the future will be for modelers to adapt or extend their theories to account for these and other regression-linked data from eye-movement studies.

## Acknowledgments

This work was supported by ESRC Grant No. RES-000-22-2069 to Don Mitchell and Tim Hodgson. Matt Green was supported by ESRC Postgraduate Studentship number PTA-031-2005-0062. We are grateful to Chuck Clifton, Keith Rayner, Erik Reichle, Tessa Warren and two anonymous reviewers for valuable comments on earlier drafts of the paper.

## Appendix A

Materials used in Experiment 1 and 2. Each sentence is constructed to take two different forms: either with the potential misanalysis region spanning Words 4–6 ("Early Misanalysis Area") or Words 7–11 ("Late Misanalysis Area"). These are abbreviated as the *Early* and *Late* forms in the list below. All misanalysis areas are marked in **bold**. In all cases, the temporary ambiguity is resolved at Word 12, which is *italicized* in each sentence. For half of the sentences, a comma was inserted immediately after the first word in the misanalysis area (which was always a past-tense verb). Experiment 1 made use only of the Early sentence forms. On half of all trials the first (and only) line-break occurred after Word 11 (marked '/'). For the other half it occurred after the disambiguating word (marked '//'). Experiment 2 used both Early and Late forms of the sentences. However, in this case there was no line-break manipulation. Line 1 always ended with Word 12 (marked '//'). Note that the proportional typeface used here provides only an imprecise indication of line length and layout for the monospaced font used for the actual experimental displays.

| | |
|---|---|
| 1-Early | While the mob **watched(,) the juggler** who was gifted and nimble / *swallowed* // a silver sword that was very sharp. |
| 1-Late | The busy guide noted that while the mob **watched(,) the juggler** *swallowed* // a silver sword that was very sharp. |
| 2-Early | While those men **hunted(,) the moose** that was sturdy and nimble / *hurried* // into the woods and took cover. |
| 2-Late | One sole hiker spotted that while those men **hunted(,) the moose** *hurried* // into the woods and took cover. |
| 3-Early | Though both lads **phoned(,) the coach** who was furious and bitter / *refused* // to permit them to join the team. |
| 3-Late | Their sad tutor moaned that though both lads **phoned(,) the coach** *refused* //to permit them to join the team. |
| 4-Early | While the baby **watched(,) her mother** who was tired and fragile / *prepared* // a new bottle of powdered milk. |
| 4-Late | The idle boy noticed that while the baby **watched(,) her mother** *prepared* // a new bottle of powdered milk. |
| 5-Early | After the vet **visited(,) the farmer** who was shifty and evasive / *admitted* // that some of his animals were ill. |
| 5-Late | The local man stated that after the vet **visited(,) the farmer** *admitted* // that some of his animals were ill. |
| 6-Early | Though the dog **sniffed(,) his trainer** who was peeved and grumpy / *avoided* // all further attempts to teach him tricks. |
| 6-Late | The alert judge mused that though the dog **sniffed(,) his trainer** *avoided* // all further attempts to teach him tricks. |
| 7-Early | While the fox **stalked(,) the geese** that were plump and healthy / *continued* // to peck at grain on the ground. |
| 7-Late | The farm hand believed that while the fox **stalked(,) the geese** *continued* // to peck at grain on the ground. |
| 8-Early | After the nun **helped(,) the refugee** who was sickly and afraid / *recovered* // slowly in the camp near the river. |
| 8-Late | The calm aide stressed that after the nun **helped(,) the refugee** *recovered* // slowly in the camp near the river. |
| 9-Early | While the maid **dressed(,) the queen** who was grouchy and aloof / *dismissed* // all the other ladies in waiting. |
| 9-Late | The high lord implied that while the maid **dressed(,) the queen** *dismissed* // all the other ladies in waiting. |
| 10-Early | After the girl **awoke(,) her father** who was drunken and drowsy / *exploded* // in anger about being disturbed so early. |
| 10-Late | The wary guest related that after the girl **awoke(,) her father** *exploded* // in anger about being disturbed so early. |
| 11-Early | After the cadet **saluted(,) the major** who was brusque and remote / *ordered* // the sergeant to prepare the ammunition. |
| 11-Late | The new NCO recorded that after the cadet **saluted(,) the major** *ordered* // the sergeant to prepare the ammunition. |
| 12-Early | After the diva **married(,) her agent** who was dynamic and astute / *secured* // her a lucrative contract with the theatre. |
| 12-Late | The nosy hack revealed that after the diva **married(,) her agent** *secured* // her a lucrative contract with the theatre. |
| 13-Early | While the team **trained(,) the striker** who was injured and unfit / *wondered* // whether the damage would take long to heal. |
| 13-Late | The news show stated that while the team **trained(,) the striker** *wondered* // whether the damage would take long to heal. |
| 14-Early | After the crowd **heckled(,) the comic** who was nervous and scared / *appeared* // to cut his act short in humiliation. |
| 14-Late | The daily rag claimed that after the crowd **heckled(,) the comic** *appeared* // to cut his act short in humiliation. |
| 15-Early | After the reps **lobbied(,) the union** that was divided and weary / *directed* // its committee to approve the proposal. |
| 15-Late | The miners all gloated that after the reps **lobbied(,) the union** *directed* // its committee to approve the proposal. |
| 16-Early | While the crew **filmed(,) the actress** who was fuming and cursing / *stormed* // off the set of the film in a tantrum. |
| 16-Late | The wise agent heard that while the crew **filmed(,) the actress** *stormed* // off the set of the film in a tantrum. |

**Appendix A** (*continued*)

| | |
|---|---|
| 17-Early | After the boxer **fought(,) the medic** who was anxious and worried / *carried* // a stretcher to the side of the ring. |
| 17-Late | The keen fan lamented that after the boxer **fought(,) the medic** *carried* // a stretcher to the side of the ring. |
| 18-Early | Though the horse **kicked(,) the trainer** who was quick and agile / *remained* // calm and managed to avoid getting hurt. |
| 18-Late | The old groom showed that though the horse **kicked(,) the trainer** *remained* // calm and managed to avoid getting hurt. |
| 19-Early | After the woman **taught(,) the pupils** who were bright and smart / *realised* // that they could now solve the equations. |
| 19-Late | The new head observed that while the woman **taught(,) the pupils** *realised* // that they could now solve the equations. |
| 20-Early | After the boss **ordered(,) the waiter** who was ancient and doddery / *mumbled* // the details to the chef incorrectly. |
| 20-Late | The grill cook remarked that after the boss **ordered(,) the waiter** *mumbled* // the details to the chef incorrectly. |
| 21-Early | While the woman **bathed(,) her husband** who was muddy and bruised / *announced* // that he wanted to have a shower. |
| 21-Late | The hotel maid joked that while the woman **bathed(,) her husband** *announced* // that he wanted to have a shower. |
| 22-Early | After the army **attacked(,) the rebels** who were quiet and swift / *launched* // a counter attack and inflicted huge losses. |
| 22-Late | The war diary argued that after the army **attacked(,) the rebels** *launched* // a counter attack and inflicted huge losses. |
| 23-Early | After the fire **burned(,) the workman** who was careful and dutiful / *laboured* // to make sure the area was secure. |
| 23-Late | The male nurse spotted that after the fire **burned(,) the workman** *laboured* // to make sure the area was secure. |
| 24-Early | While the temp **assisted(,) the tycoon** who was pompous and aloof / *committed* // a series of white collar financial crimes. |
| 24-Late | The desk clerk swore that while the temp **assisted(,) the tycoon** *committed* // a series of white collar financial crimes. |

## References


Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1998). Lexical guidance in sentence processing. *Psychonomic Bulletin and Review, 5*, 265–270.

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

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.

Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language, 44*, 297–324.

Braze, D., Shankweiler, D., Ni, W. J., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31*, 25–44.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence processing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology, 49*, 664–695.

Buswell, G. T. (1922). Fundamental reading habits: A study of their development. Supplementary Monographs No. 21. Chicago: Chicago University Press.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157–205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics in perspective. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 19–75). Westport, CT: Ablex.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 341–372). Elsevier.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Desmet, T., De Baecke, C., & Brysbaert, M. (2002). The influence of referential discourse context on modifier attachment in Dutch. *Memory & Cognition, 30*, 150–157.

Elman, J., Hare, M., & McRae, K. (2005). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Lawrence Erlbaum.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.

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

Ferretti, T. R., & McRae, K. (1999). Modeling the role of verb-bias and plausibility in the direct object/sentence complement ambiguity. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 161–166). Hillsdale NJ: Erlbaum.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalysis of garden-path sentences. *Journal of Memory and Language, 30*, 725–745.

Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology, 47*, 247–275.

Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73–100). Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Ferreira, F. (1998). *Reanalysis in sentence processing*. Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–141). Dordrecht: Kluwer Academic Publishers.

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

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

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

Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1–17.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29*, 261–290.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32*, 101–123.

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

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension—Individual-differences in working memory. *Psychological Review, 99*, 122–149.

Kennedy, A., & Murray, W. S. (1987). Spatial coordinates and reading—Comments on Monk (1985). *Quarterly Journal of Experimental Psychology, 39*, 649–656.

Kennedy, A., Brooks, R., Flynn, L.-A., & Prophet, C. (2003). The reader's spatial code. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 193–212). Oxford: Elsevier.

Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads. Evidence from eye-tracking and SRNs. In P. P. Slezak (Ed.), *Proceedings of the 4th International Conference on Cognitive Science.* (pp. 330–335). Sydney, Australia: University of New South Wales.

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

Lewis, R. L. (1993). An architecturally-based theory of human sentence comprehension. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Lewis (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247–285). Dordrecht: Kluwer Academic Publishers.

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

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35–54.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. In D. C. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing* (Vol. II, pp. 58–121). Cambridge, MA: MIT Press.

McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents of sentences. In D. E. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing* (Vol. II, pp. 272–325). Cambridge, MA: MIT Press.

McClelland, J. L., & Rumelhart, D. E. (1981). The interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review, 88*, 375–407.

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

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

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 601–681). Erlbaum.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology, 30*, 609–636.

Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye-movement analyses. *Quarterly Journal of Experimental Psychology, 40*, 693–718.

Narayanan, S., & Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading times in sentence processing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.). *Advances in neural information processing systems* (Vol. 14, pp. 59–65). Cambridge, MA: MIT Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency-effects in a multilayered perception—Implications for child language-acquisition. *Cognition, 38*, 43–102.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*, 1–56.

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

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 358–374.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word-frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*, 191–201.

Rayner, K., Juhasz, B., Ashby, J., & Clifton, C. (2003). Inhibition of saccade return in reading. *Vision Research, 43*, 1027–1034.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hove, UK: Lawrence Erlbaum.

Rayner, K., & Sereno, S. C. (1994). Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory & Cognition, 22*, 281–285.

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

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125–157.

Reichle, E. D., Warren, T., & McConnell, K. (submitted for publication). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–526.

Reilly, R. G., & Radach, R. (2003). Foundations of an interactive activation model of eye movement control in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 429–455). Amsterdam: Elsevier Science.

Rohde, D. L. T. (2002). A connectionist model of sentence comprehension and production. Unpublished Ph.D. thesis, Computer Science Department, School of Computer Science, Carnegie Mellon University.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In D. C. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing* (Vol. II, pp. 216–271). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse. Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 1521–1543.

St John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence, 46*, 217–257.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 550–569.

Stevenson, S. (1993) A competition-based explanation of syntactic attachment preferences and garden path phenomena. In *Proceedings of the 31st annual meeting of the association for computational linguistics. (pp. 266–273). Morristown, N.J: ACL.*

Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor & F. Fer-reira (Eds.), *Reanalysis in sentence processing* (pp. 327–363). Dordrecht, The Netherlands: Kluwer.

Sturt, P., & Crocker, M. W. (1998). In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 365–400). Dordrecht: Kluwer Academic Publishers.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40*, 136–150.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12*, 211–271.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: Implications for the architecture of the language comprehension system. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms of language acquisition and processing* (pp. 90–118). Cambridge, UK: Cambridge Univ. Press.

Traxler, M. J. (2005). Plausibility and verb subcategorization in temporarily ambiguous sentences: Evidence from self-paced reading. *Journal of Psycholinguistic Research, 34*, 1–30.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558–592.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning Memory and Cognition, 19*, 528–553.

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

Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285–316.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science, 32*.

Vitu, F., & McConkie, G. W. (2000). Regressive saccades and word perception in adult reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 301–326). Oxford: Elsevier.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language, 26*, 714–738.

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

Yang, S. (2006). An oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research, 7*, 56–69.