Chapter 4
Linking Eye Movements to Sentence Comprehension in Reading and Listening

JULIE E. BOLAND

Eyetracking paradigms in both written and spoken modalities are the state of the art for online behavioral investigations of language comprehension. But it is almost a misnomer to refer to the two types of paradigms by the same "eyetracking" label because they are quite different. Reading paradigms gauge local processing difficulty by measuring the participant's gaze on the very material that he or she is trying to comprehend. The critical sentence regions are determined spatially, and gaze is measured in terms of the time spent looking within a region of interest, the likelihood of a regressive eye movement out of the region, and so forth. In contrast, listening paradigms gauge how rapidly successful comprehension occurs by measuring how quickly people look, or how likely people are to look, at objects referenced by the linguistic material. In many cases, inferences can be drawn about the content of the listener's representation, based upon which of several object are fixated.

This chapter summarizes some of the contributions of each paradigm, focusing on the development of linking assumptions between eye-movement data and sentence comprehension processes. I will also discuss some limitations that currently plague each paradigm and make a few suggestions for how we might get to the next level of investigation using these paradigms. The first half of the chapter focuses on reading, and the second half on listening. It will quickly become clear that the same linking assumptions will not serve us in both reading and listening paradigms. The cognitive processes under investigation include word recognition and syntactic–semantic analysis. The behaviors that we can measure in an eye-movement paradigm are the location, duration, and onset time of each fixation. Intuitively, the linkage between cognition and fixation is straightforward in reading paradigms: Reading comprehension involves visual attention, and visual attention requires fixation. However, it is less clear, on the face of it, why eye movements should be causally linked to language comprehension in a listening paradigm.

The linking assumptions between dependent measures and mental processes are worthy of discussion in and of themselves because these linking assumptions are essentially hypotheses about one's methodology rather than assumptions of the standard type. Unfortunately, these methodological hypotheses often go unstated and untested.

0-415-93379-X/04/$0.00+$1.50

4.1 EYE MOVEMENTS AND READING

Eye-movement data from reading have been very influential for evaluating theories of human sentence processing. This is because the eye-movement record provides an on-line measure of processing difficulty with high temporal resolution, without relying on any secondary task to produce the dependent measures. Furthermore, we know much about factors that influence the planning and execution of saccades and the duration of fixations because eye movements have been carefully studied for several decades within the domains of vision, motor control, and language processing (see Rayner, 1998, for a recent review). For example, researchers have studied how our eyes move over meaningless strings and then examined how much of the variability in fixation duration and saccade landing site is linked to linguistic factors in actual text.

In fact, we have a pretty clear understanding of the amount of visual information processed during a fixation on text. During reading, factors such as a word's frequency, length, predictability, and ease of integration into the sentence influence how long it takes to access the lexical entry for that word and to incorporate the new lexical information into the structural and conceptual representations the reader is constructing for the sentence (e.g., Pollatsek & Rayner, 1990; Rayner, Sereno, & Raney, 1996). These same factors also influence whether the eyes fixate on a word and, if so, how long the fixation is maintained (Just & Carpenter, 1980; Rayner, Reichle, & Pollatsek, 1998; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Reichle, Pollatsek, Fisher, & Rayner, 1998). In contrast, we do not understand how the different dependent measures that are commonly used (see Table 4.1) are linked to specific cognitive events. For example, long reading times on a word and regressive eye movements from the current word to earlier words are both presumably caused by processing difficulty associated with the current word. But when difficulty occurs, we do not know what determines how the difficulty will manifest itself in the eye-movement record. Reichle et al.'s E-Z Reader is probably the current state of the art in terms of precise models of eye movements during reading, but it models only

AU: Citation in Table 4.3 – is it 1999a or 1999b?

Table 4.1. **Commonly Used Dependent Measures in Reading Studies**

First fixation	The duration of the first fixation in a region
First-pass time	The time spent in a region from first entering it until first leaving the region with a saccade in any direction
Regression path time	The time from first entering a region until moving the eyes beyond (rightward) the region. If there were first-pass regressions, this will include time spent in earlier regions following the regression
Probability of a regression	The percentage of regressive (leftward) eye movements out of a region; usually limited to first-pass regressions
Total time	The sum of all fixations in a region, including secondary fixations

Note: A region is usually defined as a word or phrase.

"normal" undisrupted reading. They did not even attempt to model the probability of a regressive eye movement to an earlier region of the sentence.

4.1.1 Multiple Dependent Measures: Converging and Diverging Evidence

It is undoubtedly an advantage that we can carve the eye-movement record up in different ways. Most researchers analyze three or four different dependent measures in hopes of getting a complete view of the cognitive processes involved in sentence comprehension. Oftentimes, they find the same pattern in all the dependent measures, which makes for a coherent set of results. However, I will argue that we can learn more about the linking assumptions between eye-movement data and the cognitive events from sentence comprehension by studying experiments in which the dependent measures diverge.

An Example of converging evidence is the classic paper by Frazier and Rayner (1982). Participants read garden-path sentences like those in (1). It is well known that readers and listeners develop structural (syntactic) and interpretive (semantic) representations of sentences incrementally as they read or hear each word. However, the structural position of *a mile* is temporarily ambiguous; it can either be the direct object of *jogs* as in (1a), or the subject of an embedded clause as in (1b).[1]

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

The experimenters expected readers to adopt the direct object structure and to experience processing difficulty in (1b) at the point of disambiguation, which is set in roman type. As predicted, *seems* was fixated longer and regressive eye movements occurred more frequently in (1b) than in (1a). Some trials exhibited the effect in one dependent variable, some trials in the other, and some trials exhibited the effect in both dependent variables. The linking assumption here is fairly simple: A single type of cognitive event can result in various behavioral outcomes. The combination of long fixations and regressive eye movements at the point of disambiguation has become the hallmark of a garden path in eye-movement studies of reading, and such effects can be found even when the reader is unaware of any processing difficulty.

Much more recently, Traxler, Morris, and Seely (2002) found that the difficulty of reanalysis after a garden path in object relative sentences was mitigated by animacy. This pattern was seen as a trend in the first-pass data and was reliable in their other three dependent measures (quasi-first pass,[2] the likelihood of a regression, and total reading time). Findings like these, of which there are many, suggest that processing load increases are equally likely to be reflected in longer fixations, secondary fixations in the difficult region, and regressions to earlier segments.

Although the eye-movement record can be analyzed in different ways, it is important to keep in mind that we do not have four or five dependent measures that are derived independently. Within a region of interest, there are really just two behavioral measures: fixation duration and saccade extent/direction (i.e., does the eye leave the region on the saccade, and if so, does it move forward or backward?).

PE: Slash used by author.

Everything else is computed from that. For example, the probability of a first-pass regression impacts several different dependent measures, including the probability of a regression, regression path time, and total time.

Some research questions may be more easily answered when the various dependent measures do not all exhibit the same pattern.

- Do some types of constraints influence syntactic ambiguity resolution but not the initial generation of structural alternatives?
- Which constraints determine the structures that are initially accessed or constructed and how easy it is to do so?
- Are phrase structure, morphosyntactic, and semantic operations sequentially ordered or simultaneous?

It is fairly common in reading studies to contrast early (first-pass) effects with later (second-pass or total-time) effects. In doing so, one hopes to use the multiple dependent variables offered by the eye-movement record to distinguish early lexical and structure-building processes from later processes that make use of those representations. Importantly, different researchers have taken this contrast to reflect different cognitive events, depending upon their theoretical assumptions. Some researchers have argued that detailed lexical information is not part of the initial structure-building process (e.g., Mitchell, 1987), while others have argued that detailed lexical information forms the basis of the initial structure-building process (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994). In my own research, I have argued that the architecture of sentence comprehension systems is restricted, such that lexical and syntactic constraints influence the initial construction of syntactic alternatives as each new word is integrated into the developing sentence structure. At the same time, I have maintained that constraints from higher levels of representation can influence syntactic ambiguity resolution when multiple structural alternatives are generated. The first of these two architectural claims was tested and supported by Boland and Blodgett (2001).

Boland and Blodgett (2001) embedded unambiguous target sentences like (2a) and (2b) in simple stories. The sentences contain a noun–verb homograph (*duck*), but its syntactic category is disambiguated by the preceding pronoun.

(2) Example target sentences from Boland and Blodgett (2001). Bars separate the regions used for analysis.
(a) Noun target. *She | saw his | duck and | chickens near | the barn.*
(b) Verb target. *She | saw him | duck and | stumble near | the barn.*

The examples in (2) are taken from a story about a girl visiting a boy on a farm. The sentence prior to the target sentence supported either a noun or verb meaning of the homograph: *Kate watched everything that Jimmy did*. Or *Kate looked at all of Jimmy's pets*. The type of context sentence and the type of target sentence were crossed to create four levels of discourse congruency. In addition, the relative frequency of the noun and verb forms of the homograph were varied continuously to create a lexical frequency variable. Following the architectural assumptions outlined

above, we predicted that the more frequent the appropriate form of the homograph, the easier it would be to integrate the homograph with the developing structural representation. In contrast, we expected that discourse congruency would not affect ease of structure building in this unambiguous syntactic context. Under the linking assumption in (3) then, only lexical frequency effects should be seen in the first-pass reading times of these unambiguous target sentences.

(3) **Linking assumption:** First-fixation duration reflects ease of structural integration but not pragmatic/discourse integration.

PE: Slash used by author.

As predicted, there was a lexical frequency effect in the first fixation on the region *duck and*—the earliest possible location. In contrast, no discourse congruency effect was seen in the first-pass reading times. There was a discourse congruency effect in the likelihood of regressive eye movements, but this effect was downstream, in the region after the lexical frequency effect. There was also a very robust effect of discourse congruency in the second-pass reading times, distributed throughout most of the sentence. Note that in this contrast between early and late effects, there is a difference in both the dependent variables that exhibited the effects and the sentence region where the effect was found.

The early versus late contrast in Boland and Blodgett (2001) has theoretical implications because it suggests that lexical frequency influences the initial generation of structure but discourse congruency does not. The later effects of discourse congruency are assumed to reflect anomaly detection after the sentence structure has been determined. An important part of the argument is that discourse congruency *does* have rapid and local effects on structural ambiguity resolution. In temporarily ambiguous sentences like *She saw her duck …* , Boland (1997) found effects of discourse congruency at the point of disambiguation (which was either the word following the ambiguous pronoun or downstream of the noun–verb homograph). This demonstrates that discourse congruency does resolve a syntactic ambiguity when a congruent and an incongruent structure are both syntactically possible, but that discourse congruency cannot determine what structures are possible, neither does it influence the initial structure-generating process.

PE: Author's slash changed to an en dash? OK?

4.1.2 Anomaly Detection

In Boland and Blodgett (2001), the discourse anomaly effects were observed relatively late in the eye-movement record. Under what circumstances should anomaly effects arise late? Would local structural anomalies reveal themselves earlier than discourse anomalies? To fully answer these questions, a detailed model of the mapping between the dependent measures and the underlying cognitive processes would be required. Athough no such model exists, some relevant evidence is reviewed in this section.

It should be noted that eye-movement research has not distinguished itself as much as some other methodologies in the area of anomaly detection. In particular, event-related potential (ERP) research has been quite promising in distinguishing anomaly detection processes in terms of the linguistic level of analysis at which the

anomaly occurs (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Friederici & Frisch, 2000; Gunter, Stowe, & Mulder, 1997; Hagoort, Brown, & Groothusen, 1993; Osterhout & Nicol, 1999). See van Berkum (chapter 13, this volume) and Osterhout et al. (chapter 14, this volume) for brief summaries of the N400, P600, and LAN components in semantic and syntactic anomaly detection studies. In fact, Friederici (1995) outlined a detailed processing architecture in which phrase structure construction (based on major syntactic category, as in Frazier's [1978, 1987] garden-path theory) precedes morphosyntactic and lexical-semantic processing. Her proposed architecture is motivated by differences in both the scalp location and the latency of ERP anomaly detection effects. Given the latency differences observed in the ERP record, should we not see similar distinctions in the eye-movement record, where we have more detailed temporal information?

There is surprisingly little evidence on "pure" anomaly detection in eye-movement studies. Most have focused on garden-path sentences rather than the globally anomalous sentences typically used in ERP research. A garden path occurs when a temporarily ambiguous region of a sentence is misanalyzed. In most cases, the initial analysis becomes anomalous when disambiguating words are encountered later in the sentence. For example, (1b) is a garden-path sentence in which *a mile* is structurally ambiguous and *seems* is the point of disambiguation. In some cases, the syntactically ambiguous phrase itself contains disambiguating semantic information. For example, in (4), the attachment of the prepositional phrase (PP) is ambiguous until the plausibility of using a revolver to see is evaluated against the plausibility of a cop having a revolver. Sentence (4) will cause a garden path if the reader initially attaches the PP to the verb phrase, as maintained by Rayner, Carlson, and Frazier (1983).

(4) *The spy saw the cop with a revolver, but the … .*

As noted above, garden-path effects are characterized by increased looking time in the disambiguating region and higher probabilities of regressions out of the disambiguating region (e.g., Frazier & Rayner, 1982). Because garden paths are generally taken as evidence that the language-processing system pursues only one (incorrect) analysis of the ambiguous material, the linking assumption is that the cognitive phenomena of anomaly detection and syntactic reanalysis are collectively reflected in the eye-movement data as longer first-pass reading times, an increased probability of regressive eye movements, and the time spent re-reading earlier material following a regression. In fact, Frazier and Rayner suggested more specific linking assumptions between the cognitive processes, anomaly detection and reanalysis, and the dependent measures, fixation duration and regressions, respectively. They suggested that "there was an awareness at some level on the first fixation in the disambiguating region that something was wrong, as evidenced by a longer fixation duration" (p. 193). Because regressive eye movements often landed in the ambiguous region, Frazier and Rayner also suggested that such regressions "indicate that subjects have detected an error in their initial analysis of the sentence and have identified the source of the error" (p. 203).

These particular linking assumptions between eye movements and the underlying cognitive behavior have been widely accepted over the past two decades. However, there have been some cautionary notes and some conflicting conclusions as well. For example, Rayner et al. (1989) noted that regressions "could reflect not only the existence of an erroneous initial analysis, but also the relative plausibility of competing analyses, the syntactic differences between the initial and the revised analyses, and so on" (p. 38). In fact, Spivey-Knowlton and Tanenhaus (1998) attributed increased reading times to competition between parallel structures—not to anomaly detection and reanalysis. Their constraint-based lexicalist account allows such an attribution because, in contrast to Frazier and Rayner (1982), they assume that syntactic alternatives are postulated and evaluated in parallel. In short, despite some commonly held assumptions, there are many open research questions concerning the relationships between the dependent measures in eye-movement paradigms and the cognitive processes underlying sentence processing.

If noticing something anomalous is the initial component of processing difficulty in garden-path sentences, then one would expect to see anomaly effects in the first-pass measures on the anomalous words in both garden-path sentences and globally anomalous sentences. However, a review of the small eye-movement literature on unambiguous, globally anomalous sentences suggests that this is not always true. Table 4.2 provides a summary of anomaly detection effects from the eye-movement literature. Only local first-pass measures are considered here. This includes the first-pass reading time in the anomalous region and the probability of a regression out of the region during the first pass through the anomalous region. Sometimes the regression path duration is also taken to be a first-pass measure, but it is excluded here on the grounds that it includes processing time spent in multiple regions and is therefore not a strictly local measure. The shaded areas in Table 4.2 represent the types of anomalies that were investigated in each study. Unfortunately, there have not been many papers that investigate several types of anomalies within a single experiment. This makes it difficult to determine (in analogy to the ERP experiments) whether the gaze pattern in response to an anomaly differs depending upon linguistic level of anomaly.

Boland and Blodgett (2002) examined the broadest range of anomaly types. They used two sets of critical stimuli, an "argument structure set" and an "agreement set." The first set of stimuli contained verb argument structure violations on the indirect object. Examples are given in (5) and (6). Anomalous words are starred, and doubly anomalous words receive two stars. The critical word at which the potential anomaly is apparent is underlined. The anomalous indirect object was either of the wrong phrasal category—a noun phrase (NP) instead of a prepositional phrase (PP) as in (5b)—or had the wrong semantic properties (*signs*), or both. The anomalies in the agreement stimuli were either a syntactic subject/verb agreement violation as in (6b) or a semantic violation in which the sentential subject (*snake/s*) was not a suitable agent for the verb. Thus, in both stimulus sets, the semantically anomalous words were inconsistent with the thematic role that the verb assigned to one of its arguments. However, the syntactic violations were quite different. The syntactic anomalies in the agreement stimuli were morphosyntactic agreement violations similar to those used by Ni, Fodor, Crain, and Shankweiler (1998); Braze,

PE: Slash used by author.

Table 4.2. **Anomaly Detection Results (Local First-Pass Measures)**

Experiment	Measure	Subcategory	Semantic	Morphosyntactic
Boland & Blodgett, 2002	Reading time	X		
	% Regression	X		X
Ni et al., 1998; Braze et al., in press	Reading time			
	% Regression			X
Pearlmutter et al., 1999, Exp 2	Reading time			
	% Regression			X
Deutsch & Bentin, 2001	Reading time			X
Frisson & Pickering, 1999	Reading time		X	

Shankweiler, Ni, and Palumbo (in press); and Pearlmutter, Garnsey, and Bock (1999). In the argument structure stimuli, the violations reflected a conflict between the verb's argument structure and the phrasal category of the indirect object.

(5) a. *Kim recommended Shakespeare to <u>everyone/*signs</u> after she saw Hamlet.*
b. *Kim recommended Shakespeare *<u>everyone/**signs</u> after she saw Hamlet.*

(6) a. *The canary/*snake in the large cage <u>sings</u> beautifully.*
b. *The *canaries/**snakes in the large cage <u>sings</u> beautifully.*

As noted above, there is some evidence from the ERP literature that phrasal category violations are recognized more rapidly (e.g., Friederici, Pfeifer, & Hahne, 1993), but there have been no previous attempts in the eye-movement literature to distinguish between morphosyntactic and phrasal category violations. Boland and Blodgett (2002) found that syntactic congruency affected first-pass reading time when the manipulation involved a phrasal category error, but not when it involved a morphological feature error. Both types of syntactic congruency influenced the likelihood of a first-pass regressive eye movement. Semantic congruency did not influence first-pass reading time, but it did affect the regression path duration.

Ni et al. (1998) investigated syntactic and semantic anomalies in unambiguous sentences like those in (7). Both the syntactic and semantic anomalies (illustrated in (7a) and (7b), respectively) led to more regressive eye movements than the control sentences (7c). The likelihood of an immediate regression, however, was higher for syntactic anomalies than for semantic anomalies. In contrast, only the semantic anomaly induced longer first-pass fixations, and only after (rather than during) the anomalous region. Ni et al. interpreted these results to suggest qualitative differences in the cognitive response to the two types of anomalies—differences that are directly reflected in the eye-movement patterns.

(7) a. **It seems that the cats won't usually <u>eating</u> the food …*
b. **It seems that the cats won't usually <u>bake</u> the food …*
c. *It seems that the cats won't usually <u>eat</u> the food …*

More recently, Braze et al. (in press) focused on regressive eye movements in comparing morphosyntactic and semantic anomalies. They found that syntactic anomalies elicited an immediate peak in the percentage of regressive eye movements, while semantic anomalies led to a gradual rise in regressions that peaked at the end of the sentence.

PE: Author's slash changed to en dash.

Pearlmutter et al. (1999, Experiment 2) examined subject–verb agreement errors in sentences like (8). As in Ni et al. (1998) and Braze et al. (in press), the eye-movement pattern was dominated by regressions. However, Pearlmutter et al. *did* find longer first-pass fixations for some conditions on a subset of trials. The effect was not localized to the anomalous word alone (*were*), but became apparent when fixations on the anomalous word and the following word were summed. Furthermore, the effect was only observed after excluding the trials on which there was a regression out of that region (about 14% of trials) and excluding eight participants who either always or never made regressions out of that region.[3] I have not included the Pearlmutter et al. potential reading-time effect in Table 4.2 because it was not strictly localized to the anomalous word.

(8) a. *The key to the cabinet* <u>*was/*were*</u> *rusty …*
b. *The key to the cabinets* <u>*was/*were*</u> *rusty …*

PE: Slash changed to en dash.

Deutsch and Bentin (2001) examined subject–verb gender agreement in Hebrew, a language that marks both gender and number in the verb morphology. In contrast to the English morphosyntactic results, they found an anomaly effect for mismatched gender in sentences like (9a) in the first-pass reading times on the verb. In (9a), the plural verb explicitly marks the gender of the subject. Only second-pass reading effects were observed for the unmarked (singular) form (9b). The authors did not analyze the probability of regressive eye movements. The authors also reported an ERP version of the experiment. They found a negative component 80–250 ms after incongruent verbs, regardless of markedness, and a P600 for marked incongruent verbs only.

(9) a. *I enjoyed seeing how the actors (*hasaxkani**ot**/hasaxkani**im**)* <u>*were enchanting (maksim**im**)*</u> *…*
b. *The woman saw that the boy/girl (*hayeled/hayalda)* <u>*had fallen (nepal)*</u>
…

Frisson and Pickering (1999) reported some early and some late semantic anomaly effects in their investigation of metonymic expressions. Familiar metonomy (*Americans protested during* Vietnam.) is handled easily; the authors report that readers were able to coerce the place (Vietnam) into an event (the Vietnam war) with no increase in processing difficulty over a literal control (*Americans hitchhiked around Vietnam.*). However, unfamiliar metonymic expressions seem to be treated as semantic anomalies, as in (10a), *Finland* condition. As in the Boland and Blodgett experiment, there was no evidence of an anomaly in the first-pass reading times for items like (10a). In contrast, Frisson and Pickering reported local first-pass reading time effects for semantic anomalies like those in (10b). The crucial difference

between (10a) and (10b) is unclear. The authors suggest that *Finland* in (10a) initiates a search for relevant events, while there is a clear feature clash or selectional restriction in (10b). However, the details of such an account need to be carefully worked out. Why should unfamiliar "place-for-institution" metonomy violate a selectional restriction if unfamiliar "place-for-event" metonomy does not?

(10) a. *A lot of Americans protested during* *Finland/Vietnam …
b. *The blasphemous woman had to answer to the* *stadium/convent …

What generalizations can we make about these anomaly effects in unambiguous sentences? Can we generate any predictions about the types of anomalies that will generate local first-pass effects? There was only one study that examined phrasal category violations, and unfortunately, it is unpublished (Boland & Blodgett, 2002). Eye-movement evidence of the violations emerged in the first-pass reading times on the anomalous material. In fact, the results are predicted by most current sentence-processing theories, which maintain that verb subcategorization guides the initial structural analysis. However, not all syntactic violations led to local first-pass reading time effects. There was much more data on morphosyntactic errors, and within this group, most of the studies examined subject–verb agreement violations. The English studies consistently found morphosyntactic effects in the probability of a first-pass regression, while they did not find increased first-pass reading times on the anomalous word (Boland & Blodgett, 2002; Ni et al., 1998; Braze et al., in press; Pearlmutter et al., 1999). In contrast, a Hebrew study did find first-pass reading time effects for some of the anomalous conditions—they did not analyze the probability of a regression (Deutsch & Bentin, 2001). The results regarding semantic/pragmatic violations are also mixed. Several studies found no local first-pass effects (Boland & Blodgett; Braze et al.; Ni et al.), but Frisson and Pickering (1999) did find first-pass reading time effects for certain semantic violations.

PE: Slash changed to en dash.

PE: Slash used by author.

Based upon the anomaly detection results summarized in Table 4.2 as well as data from grammatical stimuli such as that in Boland and Blodgett (2001), I offer the linking assumption in (11) as a working hypothesis. "Structural integration" is intended here to mean adding the new constituent to the developing phrase-structure tree. When trouble arises, I assume that the reader will either maintain gaze or refixate the problematic word until either structural integration is complete or has been determined impossible. The English morphosyntactic violations represent an example in which structural integration proceeds normally but the agreement violation triggers reprocessing of earlier material.

(11) **Linking Assumption:** The eyes do not leave a word until it has been structurally integrated. Therefore, constraints that control structure-building influence first-pass reading time.

The structure-building constraints that influence first-pass reading time would include syntactic category and subcategory, lexical frequency, morphological agreement in richly case-marked languages (like Hebrew, but not English), and semantics when it determines structure.[4] Note that the relevant constraints proposed

here are quite different from the traditional first-pass assumptions by Frazier and colleagues in garden-path literature (e.g., Frazier, 1978, 1987). In Frazier's garden-path model, neither detailed lexical information (such as verb subcategorization and the relative frequency of noun–verb forms of a homograph) nor semantics can influence the initial rule-generated parse.

PE: Slash changed to en dash.

The proposal that eye fixations are related to structural integration may seem to be quite a departure from the current state-of-the-art models of eye movements such as E-Z Reader (Reichle et al., 1998). There are at least two issues here. The first issue is the assumed relationship between lexical access completion and syntactic analysis. E-Z Reader assumes the traditional distinction between lexical and syntactic processing embodied in Frazier's garden-path model. This sharp distinction allows E-Z Reader to predict eye movements based upon atomic lexical variables alone, such as the familiarity of the word in isolation, ignoring syntactic and high-level text factors. While this strategy simplifies the modeling task, it is at odds with the increasingly lexicalized nature of syntactic parsing theories (e.g., Ford, Bresnan, & Kaplan, 1982; Carlson & Tanenhaus, 1988; McElree, 1993). In many current approaches, word recognition within sentence context cannot be separated from structure assignment (e.g., Boland, 1997; Lewis, 2000; MacDonald et al., 1994; Novick, Kim, & Trueswell, 2003). That is, the identification of a single lexical form simultaneously completes word recognition and determines the local syntactic structure some or all of the time. Under such a view, the proposal that fixations are causally linked to structural integration is not a violation of E-Z Reader's core assumption that visual attention moves to the next word when lexical access of the current word is complete. Rather, such a view would require E-Z Reader to consider combinatory factors in addition to atomic lexical factors in predicting the time course of lexical access.

The second issue is the coupling of visual attention and fixation location. As in the linking assumption from (11), early theories of eye movements made a locality assumption, such that visual attention and eye fixation were tightly linked (Just & Carpenter, 1980; Morrison, 1984). In Morrison's theory, completion of lexical access on word n simultaneously prompted a shift in attention to word $n + 1$ and the planning of an eye movement to word $n + 1$. At some level of abstraction, a close relationship between visual attention and fixation location is absolutely necessary to explain reading behavior. However at a more precise level, the Morrison assumption is too constraining because it leaves us unable to model preview and spillover effects as well as refixations on the same word. To address these limitations, Reichle et al. (1998) decoupled visual attention shifts from saccade planning by assuming that there are two relevant events during lexical access. The first is completion of a familiarity check on the currently attended word, which signals an eye-movement program for a saccade to the next word. The second is completion of lexical access, which shifts attention to the next word. Thus, in apparent contrast to the linking assumption in (11), Reichle et al. do not assume that either structural integration or lexical access has been completed prior to eye-movement planning. Rather, it is the familiarity check that provides the important signal to the motor system.

To be clear, the proposed familiarity check is not an event that can be found in current theories of visual word recognition. Reichle et al. (1998) note that "The division of lexical access into two discrete serial stages, f (familiarity check) and lc (completion of lexical access), is largely a modeling convenience." (p. 133, parentheticals added here for clarity). To model the time course of the familiarity check, Reichle et al. used log word frequency (Kucera & Francis, 1982) and predictability (essentially the cloze value of each word in a sentence), two variables that robustly predict lexical decision time, a common measure of lexical access. In other words, the function of the familiarity check is to estimate when lexical access will be complete, so that visual attention and fixation location will remain closely, but not exactly, synchronized. (In E-Z Reader, lexical access completion is more strongly affected than the familiarity check by the fixation location parameter and is assumed to require additional, but unspecified, processing beyond the familiarity check.)

AU: Not included in ref. list.

To create a version of E-Z Reader that is consistent with the linking assumption in (11), factors that predict structural integration would have to be incorporated when modeling the time course of the "familiarity check." Unfortunately, the best way to do so is theory-dependent because the factors that affect structural integration are still hotly debated in the parsing literature. Nonetheless, some examples of the required changes are (a) to measure predictability in terms of the syntactic (and possibly semantic) categories of words, in addition to measuring the predictability of specific lexical items, and (b) to use the frequency of the required syntactic form instead of the basic word frequency.

Of course, even if the proposed linking assumption in (11) is correct, it leaves much to be resolved. For example, what is the relationship between regressive eye movements out of a region and long/repeated fixations within that region? Why should first-pass reading time be linked to initial structure generation rather than complete understanding? The proposed linking assumption implies that only structure-determining factors will increase first-pass time, but there are obvious exceptions to that generalization, such as the discourse level effects reported by O'Brien, Shank, Myers, and Rayner (1988) and Garrod, Freudenthal, and Boyle (1994). If the proposed linking assumption is to be maintained, one must identify the conditions that govern how quickly such discourse level effects arise.

PE: Slash used by author.

4.1.3 *The Promise and Limitations of Reading Paradigms*

Most of eye-movement research in reading has focused on garden-path sentences, leading to many insights about syntactic ambiguity resolution. However, because garden paths involve both anomaly detection and reanalysis, it is difficult to establish the mapping between cognitive events and eye-movement behaviors. In this chapter, I have focused on eye-movement patterns over *unambiguous* sentences as a means to contrast constraints on initial syntactic generation with constraints on other processes, such as syntactic ambiguity resolution, morphological agreement (in languages like English), and semantic-pragmatic felicity. I have argued that first-fixation and first-pass reading times in unambiguous sentences may provide a relatively pure index of syntactic generation. Thus, first-pass duration is influenced by lexical frequency but not discourse congruency (Boland & Blodgett, 2001),

PE: Slash used by author.

consistent with an architecture in which structure generation is largely lexical, with discourse constraints operating upon the lexically generated structures. Likewise, within anomaly detection paradigms, first-pass duration is influenced by (sub)category and other structure-determining constraints.

Despite some measure of promise, it should be clear that there has been no definitive mapping between specific cognitive processes and eye behavior. Many questions remain unanswered. Do (specific) cognitive events trigger a saccade? Is there some cognitive event that causes the eyes to regress rather than move forward? How much cognitive processing occurs in the parafovea? The answer to this last question is linked to another set of questions: Is attention limited to the currently fixated word? Do attention and gaze always shift in unison? To what degree are words recognized and structured in parallel? If there is some degree of parallelism, should we still think of the first fixation on a word as the earliest possible measure of processing for that word? Are nonfixations on a word cognitively meaningful?

In contrast to the ERP literature, there is no evidence that syntactic anomalies disrupt the eye-movement record in one way and semantic anomalies disrupt it in another. We may always have to rely upon converging evidence from ERP or MEG to understand which levels of representation have registered an anomaly. Although eyetracking paradigms lack precision in addressing that kind of "how" question, they excel at answering the "when" questions, providing a detailed temporal record of local processing difficulty.

An inherent limitation is actually the reading itself. When relying upon reading as an intermediary skill in order to measure language comprehension, one immediately limits one's population to that of skilled readers (omitting children, language-disordered populations, etc.). Furthermore, many interesting questions about language comprehension can only be answered within listening paradigms and by "situating" language within a real-world context. In the limit, one would like to study language comprehension in conversation, its original and most natural context. To address these kinds of concerns, spoken language paradigms have a great deal of appeal.

AU: What does this mean?

4.2. EYE MOVEMENTS AND LISTENING

Over the last 5 to 10 years, psycholinguists have been exploiting Cooper's (1974) finding that we tend to look at things as they are mentioned. Cooper used a passive listening task; he presented a set of pictures in a grid and found that listeners' looks to objects were time-locked to mention of those objects in a story. In more recent studies, some researchers have made the link between mentioning an object and looking at it more explicit by asking listeners to move or point to objects within a directed action paradigm. Regardless of the task (directed action versus passive listening), eye movements in listening paradigms do not provide the multilayered dependent measures found in reading paradigms (as in Table 4.1) because the looks of interest are not integrated within some regular sequence of eye movements as they are in reading. In listening paradigms, the dependent measures are usually

limited to the probability and duration of a look to a relevant object within some temporal interval.

In reading research, a natural starting point was to study how the eyes move over meaningless strings and then examine how the variability in fixations and saccades is associated with linguistic factors in actual text. No comparable approach has been taken in the listening literature, though one could—and perhaps ought to—in the struggle to outline the linking assumptions between visual attention and listening comprehension. For now, the most we can do is look to the substantial literature on scene perception in order to ground our predictions about when and how eye movements should occur.

Research on scene perception has established that scenes are identified within the first fixation (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce & Pollatsek, 1992; Hollingworth & Henderson, 1998). That is, basic global information about the scene is absorbed within a single fixation. The mean fixation duration in scene perception is about 330 ms, with a mode of 230 ms (Henderson & Hollingworth, 1999). The initial fixation patterns are quite similar across participants, but variance increases rapidly over time (Mannan, Ruddock, & Wooding, 1995). The locations of the initial fixation on a scene are based on visual, not semantic, features (Henderson & Hollingworth). However, the likelihood of refixation is based upon task-dependent informativeness (Henderson & Hollingworth). Finally, our working memory representation of scene is abstract/conceptual rather than strictly visual (Henderson & Hollingworth).

Chedk spacing of lines here please.

PE: Slash used by author.

The scene perception literature suggests that gaze during the first second or so of looking at a scene is driven by visual parameters, with little variance in performance.[5] In language comprehension experiments, the time interval of interest generally follows scene presentation by several seconds, so we should expect listeners to have already scanned the scene and have an abstract representation of it. This mental representation allows listeners to look at maximally informative regions of the scene in response to linguistic input. In fact, the Cooper (1974) results suggest a simple linking assumption: *The probability of looking at an object increases when the object is mentioned.* This linking assumption has ecological validity in conversation because we need to know properties of referents beyond their linguistic label. Directed action tasks further increase the probability of a look by using eye–hand coordination to mediate comprehension and eye movements. That is, it is difficult to manipulate or point to an object without first looking at it.

In short, listeners have reason to look at referents as they are mentioned in both passive listening and directed action tasks. This phenomenon has proven useful and informative in a variety of domains, as illustrated in Table 4.3. Research on spoken language comprehension has shown that eye fixations are time-locked to lexical access of isolated words (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), identification of referents for syntactically ambiguous phrases (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and pronoun resolution in discourse context (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). Eyetracking experiments investigating language comprehension have been conducted on children (Trueswell, Sekerina, Hill, & Logrip, 1999) and in the context of conversation among adults (Brown-Schmidt, Campana, & Tanenhaus, 2002).

AU: 1999a or 1999b?

Table 4.3. **Important Contributions From Listening Paradigms**

Cohort competition and frequency effects in lexical access	Allopenna, Magnuson, and Tanenhaus, 1998; Dehan, Magnuson, and Tanenhaus, 2001
Incremental reference assignment	Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy, J. C., and Tanenhaus, 1995
Visual referential context influences ambiguity resolution	Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995
Young children do not use visual context as efficiently in ambiguity resolution	Trueswell, Sekerina, Hill, and Logrip, 1999
Initial forays into conversational interaction	Brown-Schmidt, Campana, and Tanenhaus, 2002

Clearly, measuring eye movements in a listening paradigm can provide a sensitive index of referential success or ambiguity resolution once linguistic input has been comprehended. However, it is not obvious that eye fixations are time-locked to developing syntactic and semantic representations in a manner that can be distinguished from the listener's ultimate conceptual representation of the linguistic input. Can this research paradigm be used to investigate how people develop linguistic representations of spoken language input? Can we use it to unpack the cognitive operations in syntactic processing? Is it on-line enough to investigate parsing?

4.2.1. Anticipatory Looks

Intriguingly, reference resolution is sometimes completed prior to actual mention of the referent. For example, Sedivy, Tanenhaus, Chambers, & Carlson (1999) found that when listeners were asked to *Pick up the tall glass …* , they often planned an eye movement to the glass during the adjective, prior to hearing *glass*. Instead of waiting for bottom-up evidence that the target object is being mentioned, listeners used the current visual context (i.e., presence of a tall/short contrast set) to select the pragmatically appropriate referent. In a more recent study, Sussman, Campana, Tanenhaus, and Carlson (2002) found that listeners made an eye movement to an appropriate instrument (a pencil) when hearing *Poke the dolphin* but not *Touch the dolphin*. Thus, even though no instrument was mentioned, listeners used their knowledge about the two verbs to decide whether to manipulate the dolphin with their finger or a pencil in the real-world environment. Listeners in the Sussman et al. study were also sensitive to contextual factors that altered verb meaning. For example, they looked at a potato peeler when asked to *Peel the potato*, but not when asked to *Peel the banana*.

PE: Pl. check the levels of this head and other heads. The author's arrangement is not clear.

4.2.2 Does Argument Structure Implicitly Introduce New Entities Into Discourse?

Anticipatory looks such as those in the foregoing discussion may provide clues to intermediate representations. For example, Sussman et al. (2002) concluded that hearing a verb provides access to its thematic grids and listeners then use context to select the appropriate grid. If the relevant thematic grid contained an instrument, eye movements were observed to an appropriate instrument in the environment. A related phenomenon has been reported in reading studies (Mauner, Tanenhaus, &

Carlson, 1995, implicit agents; Carlson & Tanenhaus, 1988, open thematic roles). Just as in Sussman et al., it was argued that thematic role information from verb argument structure was accessed during word recognition. In Mauner et al.'s reading studies, the thematic roles guided comprehension even if an argument was not explicitly mentioned. Together, these studies suggest that the thematic roles accessed during verb recognition can be used both to interpret current discourse entities and to introduce new entities into the discourse.

Of course, directed action tasks such as that used by Sedivy et al. (1999) and Sussman et al. (2002) give rise to the concern that listeners are strategically guessing the speaker's intent rather than allowing language processing to proceed automatically. Arguably, normal conversation involves a great deal of strategic guessing about the speaker's intent, so this is not a problem if the goal is to study the complete comprehension process. However, if there are some partially or fully automatized aspects of syntactic and semantic processing, the directed action paradigm is not ideal for studying the representations that result from those automatized processes alone. For example, one might question whether the recognition of "poke" *obligatorally* introduces an instrument into the discourse model. And importantly, are the discourse elements that can be introduced by the verb limited to members of its thematic grids? (See Koenig, Mauner, & Bienvenue, in press, for one current approach to understanding how semantic participants are lexically encoded.) In other words, do the verb's arguments hold a privileged status or are all related words and concepts accessed in the same way?

AU: Is this word correct?

Encouragingly, work from Gerry Altmann's lab, as well as some of my own recent work, provide converging evidence for the automatic activation of thematic role information from passive listening tasks. In an extremely interesting study, Altmann and Kamide (1999) had people listen to a sentence like *The boy will move/eat the cake* while looking at a semirealistic scene with a boy, a cake, and other movable but not edible objects. Altmann and Kamide found faster looks to the cake following *eat* compared to *move*. In fact, participants often looked at *cake* in the *eat* condition prior to the onset of the noun. Altmann and Kamide concluded that the verb's thematic roles were used to proactively restrict the domain of subsequent reference.

Even in a passive listening task, it is difficult to identify the cause of the anticipatory fixations on the cake because both linguistic and general-world knowledge could have contributed to the effect. If it is solely the verb's argument structure that is driving eye movements, then listeners should not look at a bed upon hearing *The girl slept* because *bed* cannot be an argument of *slept*. However, beds are part of a prototypical sleeping event and are thus conceptually related to *sleep*. Furthermore, discussions about sleep often include mention of a bed, so linguistic co-occurrence frequency is high and the co-occurrence of sleeping and beds in participants' actual experience is likely to be extremely high. One might consider an account of Altmann and Kamide's (1999) effect that is akin to semantic priming—a conceptual, essentially intralexical, process. However, in more recent work, Kamide, Altmann, and Haywood (2003) found that combinatory semantics rather than simple lexical relationships influenced eye movements. For example, when viewing a carnival scene, listeners looked at a motorcycle upon hearing *The*

man rode … and looked at a merry-go-round upon hearing *The girl rode*. … Thus, something higher-level than simple lexical associations influenced the pattern of eye movements.

Using a similar paradigm, Boland (2003) investigated the hypothesis that the use of a verb would implicitly introduce relevant entities (linguistic arguments) that had not yet been mentioned, and thus a picture corresponding to such an entity would draw anticipatory looks. For example, upon hearing … *mother suggested* …, participants would look at a potential recipient of the suggestion. The first experiment manipulated both the argument structure of the verb and the typicality–co-occurrence frequency of the target argument or adjunct, in order to distinguish between anticipatory looks to arguments specifically and anticipatory looks to pictures that were strongly associated with the verb but did not have the linguistic status of argument. Example stimuli are in (13). The intransitive-location stimuli provide a clear case of an adjunct target (*bed/bus*), the dative-recipient stimuli provide a clear case of an argument target (*teenager/toddler*), and the action-instrument stimuli provide an intermediate case in which the targets are arguably adjuncts (*stick/hat*). Acceptability ratings insured that sentences with typical targets were judged more acceptable than sentences with atypical targets. Furthermore, typical targets were more likely to co-occur with their verbs.[6] Importantly, there was no evidence that typical recipients had a higher co-occurrence frequency than typical locations—if anything, the opposite was true.

(13) Example stimuli from Boland (2003), Experiment 1. The typical/atypical target is.underlined.
   a. Intransitive-Location: *The girl slept for a while on the bed/bus this afternoon.* (pictures: girl, bed/bus, pillow, toy car)
   b. Action-Instrument: *The donkey would not move, so the farmer beat it vigorously with a stick/hat every day.* (pictures: donkey, farmer, stick/hat, grass)
   c. Dative-Recipient: *The newspaper was difficult to read, but the mother suggested it anyway to her teenager/toddler last week.* (pictures: newspaper, mother, teen/toddler, dictionary)

The primary finding was that dative verbs prompted more anticipatory looks to potential recipients than transitive action verbs prompted to potential instruments or intransitive verbs prompted to potential locations. This argument status effect began about 500 ms after verb onset, suggesting that it occurred soon after lexical access of the verb. If verbs specify the syntactic and semantic constraints on their arguments, recognizing a verb would make available knowledge about that verb's arguments, and likely candidates (entities that satisfy the syntactic and semantic constraints) could be identified in the current discourse model or the situational context. No overall typicality effects were found, in contrast with Kamide et al. (2003). This apparent discrepancy may be due to a difference in the way the visual stimuli were presented. In Kamide et al., argument structure knowledge introduced an abstract ridable object into the discourse. In the visual scene, there were two ridable objects (a motorcycle and a merry-go-round), so real-world knowledge

guided the viewer to the most plausible one. In Boland's experiment, the argument structure of the dative verbs introduced an abstract recipient, but there was only one potential referent pictured. In both the typical and atypical conditions, the potential referent met the lexical constraints on recipients for that particular verb. Thus, plausibility had no opportunity to play a role if we assume that verb argument structure is first used to identify possible referents. (If verb argument structure and plausibility simultaneously and jointly constrain referent identification, then typicality effects would be expected even if only one possible referent were pictured.) This explanation is consistent with prior findings that pragmatic constraints influence ambiguity resolution but not the generation of linguistic structure (Boland, 1997).

Boland (2003) reports a follow-up experiment that supports this explanation of the discrepancy with Kamide et al. (2003). For some participants, the sentences from the first experiment were used again, but both typical and atypical targets were pictured on each trial. Another group of participants viewed the slides while listening to music, instead of the sentences, to establish baseline looking rates for each of the pictures. The argument status effect from the first experiment was replicated, even after a general bias to look at the recipient pictures was taken into account. A typicality effect comparable in size to those of Kamide et al. also emerged for recipients, but not for instruments or locations.

The argument status effect was replicated again in Boland's (2003) third experiment, in which a single animate NP (and the corresponding picture) served as an argument in the dative condition (14a) and as an adjunct in the action verb condition (14b). No instrument was mentioned in the critical trials, though a prototypical instrument for the action verb was always pictured and, in filler trials, pictured instruments were mentioned. There were no reliable differences in co-occurrence frequency among the dative-recipient, action–benefactor, and action–instrument pairs.

(14) *One window was broken, so the handyman …* (pictures: window, handyman, couple, tools)
   a. *mentioned* ***it right away to the owners.*** (Recipient-Argument)
   b. *fixed* ***it hurriedly for the owners.*** (Benefactor-Adjunct)

As in the first two experiments, there were more looks to the target picture when it was an argument (recipient) than when it was an adjunct (benefactor, instrument) during the interval 500–1000 ms after the onset of the verb. There was no difference in the probability of a look to a prototypical adjunct (*fix*-tools) and an improbable adjunct (*mention*-tools). Thus, the results from all three experiments indicate that linguistic constraints play a privileged role in guiding visual attention in this passive listening paradigm. Co-occurrence frequency does not provide an alternative explanation. These argument status effects suggest an important distinction between adjuncts and arguments in terms of how verbs introduce associated entities into the discourse.

4.2.3 *Limitations and Open Questions in Listening Paradigms*

Listening paradigms provide an exciting opportunity to investigate spoken language comprehension in an on-line manner. While the paradigm is largely limited to referential processing, it has been used to address a wide range of questions in word recognition, sentence analysis, and discourse processing. Because listening paradigms do not require reading, they offer the opportunity to study language processing in children and other populations that include less skilled readers. Perhaps most importantly, listening paradigms allow, at least in principle, for investigations of language processing within conversational contexts.

Despite some clear advantages over reading paradigms, listening paradigms share the lack of an explicit model linking sentence comprehension to eye movements. In fact, this problem seems worse within listening paradigms, which have not benefited from the same degree of scientific scrutiny over the past several decades. On the bright side, we are likely to make some progress if we examine and test our linking assumptions in a rigorous fashion. Some of the open questions include the following:

- How do directed action and passive listening tasks differ? Does the task influence the eye movements above and beyond the linguistic material?
- What cognitive events prompt a saccade to an object? How does this interact with task, local syntactic ambiguity, and other factors?
- Can we integrate models of scene perception to explain more variance in passive listening paradigms?
- How can we measure the likelihood that an object will be mentioned? How do we decide the appropriate label for an object when measuring likelihood?
- Are all of the effects essentially referential—that is, at the level of the discourse model? Could we find a syntactic complexity effect? Could we find a lexical bias effect (e.g., ease of integrating noun–verb homophones)?

PE: Slah changed to en dash.

4.3 CONSIDERING READING AND LISTENING TOGETHER

Thus far, I have considered reading and listening eye-movement paradigms separately for reasons outlined in the introduction: the two approaches are very different. To the extent that they measure comparable aspects of language comprehension, the two classes of paradigms might provide converging evidence for particular processing models or phenomena. In fact, there is growing evidence of this type of convergence in the literature.

One example of complementary reading and listening data is the argument–adjunct line of research from my own laboratory. Using the stop-making-sense task (Boland & Boehm-Jernigan, 1998), word-by- word reading (Blodgett & Boland, in press), and eyetracking (Boland & Lewis, 2001), reading studies have found consistent support for the view that argument structures largely determine the

strength with which competing analyses are initially made available. For example, Boland and Boehm-Jernigan investigated temporarily ambiguous dative sentences like those in (15). Following a theme-NP, like *the letter*, *give* licenses a recipient argument beginning with the preposition *to*. This argument structure constraint predicts a garden path in (15b), but not in (15c) or (15d). However, there is another possibly relevant constraint that is not related to argument structure: the prepositions *to*, *for*, and *about* differ in the likelihood of verb phrase (VP) attachment. Making a VP attachment in either case would render the recipient in PP2 ungrammatical (or at least very awkward) because an adjunct cannot precede an argument. In fact, the argument structure constraint produced the expected garden-path effect at *his*, with only (15b) differing from the control condition, indicating the primacy of the argument structure constraint. In addition, there was evidence of a garden path in (15c) one word later, indicating that the preposition's attachment bias is stored in the lexicon and plays some role during syntactic ambiguity resolution.

(15) *Which friend did John give a letter …*
  a. *… to __ for his son a month ago?* (early gap control condition)
  b. *… to his son to__ a month ago?* (garden-path condition)
  c. *for his son to __ a month ago?* ("for" is equibiased for NP and VP attachment)
  d. *about his son to __ a month ago?* ("about" has strong NP attachment bias)

In other respects, convergence has not yet been achieved, but efforts toward that goal are likely to lead to some important insights. For example, I have suggested that the text-based statistics commonly used to predict reading time in recent constraint-based lexicalist studies are inappropriate for evaluating anticipatory looks in listening studies. Compare the reading study of McDonald and Shillcock (chapter 5, this volume) to the Boland (2003) listening study described in the preceding text. McDonald and Shillcock found that the higher the transitional probability of a noun, given a verb, the shorter the first fixation duration on that noun when it followed that verb in a sentence. Because they were measuring fixations on a linguistic object (a printed noun), they measured the transitional probabilities of that object, given another linguistic object (a printed verb) in text corpora. The assumption is that both the transitional probability and the fixation duration are directly related to the time it takes to access and integrate that noun. In contrast, Boland measured looks to an unlabeled picture following a verb in a spoken sentence. In such a case, probability may play some role in guiding eye movements, but it is not clear that the relevant probabilities can be obtained from a text corpus. The dependent measure assumes that the verb has been recognized, but it does not require accessing and integrating a specific word or lexical category associated with the picture because the relevant words have not yet been uttered. Rather, the dependent measure probably reflects attention to some conceptual domain that has been made salient by the pictured material and made relevant by the linguistic input thus far. Whatever the appropriate measure of probability is, it should take into account the visual display as well as the linguistic material.

My point in comparing McDonald and Shillcock (chapter 5, this volume) with Boland (2003) is this: In reading studies, we can account for a substantial portion

of the variance in reading time by assuming that the linguistic context is the entire context, for all practical purposes. Furthermore, we generally assume that syntactic and semantic processing of some portion of text begins sometime after that text is initially fixated. However, the listening paradigm forces us to wrestle with aspects of syntactic and semantic processing that begin prior to bottom-up word recognition, and with the fact that most real-life contexts are considerably richer than the linguistic context alone. While these additional layers of complexity complicate the experimental logic, they also bring some degree of real-world validity to the experiment. And it is likely that many of the insights to be gained from the listening studies will be relevant for reading studies as well.

4.4 SUMMARY

Eye-movement data in both visual and auditory modalities have excelled in addressing "when" questions, such as how early during comprehension we have access to lexical information or how early a particular constraint is used. In this chapter, I highlighted some promising findings in each modality that push the linking assumptions between eye and mind a little further than the current state of affairs. First I suggested that, when reading unambiguous sentences, the first-fixation or first-pass reading time may be closely tied to syntactic generation (the access to, or construction of, possible syntactic structures). Second, I offered evidence that argument-structure knowledge allows us to focus on relevant entities as soon as spoken verbs are recognized, based on anticipatory looks to objects that might be event participants. Nonetheless, neither reading nor listening paradigms may be capable of distinguishing among linguistic (e.g., syntactic vs. semantic) levels of representation, unlike ERPs which have proven sensitive to syntactic–semantic distinctions.

4.5 ACKNOWLEDGMENTS

This chapter is based upon a presentation to the 8th Annual AMLaP conference, held in Tenerife, Canary Islands, in September, 2002. I would like to thank the members of the audience and the other presenters for their comments. As I was preparing for the talk and writing this chapter, I benefited from the contributions of my collaborators, students, and colleagues, especially Allison Blodgett, David Thomas, Meghan Ahern, Alicia Seifers, and the University of Michigan Psycholinguistics Research Group.

References

Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language, 38*(1), 112–130.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Altmann, G. T. M. (1994). Regression-contingent analyses of eye movements during sentence processing: Reply to Rayner and Sereno. *Memory and Cognition, 22,* 286–290.

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., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition.*

AU: Volume number?

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergrouping relational violations. *Cognitive Psychology, 14,* 143–177.

Blodgett, A., & Boland, J. E. (in press). Differences in the timing of implausibility detection for recipient and instrument prepositional phrases. *Journal of Psycholinguistic Research.*

AU: Published yet?

AU: Volume number?

Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes, 12,* 423–484.

Boland, J. E. (2003). *Visual arguments.* Manuscript submitted for publication.

AU: Published yet?

Boland, J. E., & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language, 39,* 684–719.

Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language, 45*(3), 391–411.

Boland, J. E., & Blodgett, A. (2002). *Eye movements as a measure of syntactic and semantic incongruity in unambiguous sentences.* Unpublished manuscript, University of Michigan at Ann Arbor.

Boland, J., & Lewis, R. (2001). *Distinguishing generation and selection of modifier attachments: Implications for lexicalized parsing and competition models.* Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 531–543.

AU: Published yet?

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (in press). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research.*

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). *Reference resolution in the wild: How addresses circumscribe referential domains in real time comprehension during interactive problem solving.* Paper presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York.

Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and Semantics: Vol. 21. Thematic Relations*. San Diego, CA: Academic Press.

Cooper, R. M. (1974). Control of eye fixation by meaning of spoken language—new methodology for real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317–367.

AU: Spelled Dehan in text citation.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*(2), 200–224.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409–436.

Ford, M., Bresnan, J., & Kaplan, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

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

Frazier, L. (1987). Theories of sentence processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding.* (pp. 291–307). Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50,* 259–281.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43,* 476–507.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research, 1*(3), 183–192.

Frisson, S. P., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1366–1383.

Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the online resolution of sentencs in a discourse. *Journal of Memory and Language, 33,* 39–68.

Griffen, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Gunter, T., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34,* 660–676.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439–483.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–270.

Hollingworth, A., & Henderson, J. (1998). Does consistent scene context facilitate object perception? *Journal of Experiment Psychology: General, 127,* 398–415.

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

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language.*

Koenig, J.-P., Mauner, G., & Bienvenue, B. (in press). Arguments for adjuncts. *Cognition.*

AU: Please provide volume number.

AU: Published yet?

Lewis, R. L. (2000). Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In M. W. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing.* New York: Cambridge University Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D visual images. *Spatial Vision, 9,* 363–386.

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language, 34,* 357–382.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language, 32,* 536–571.

Mitchell, D. C. (1987). Lexical guidance in human parsing—locus and processing characteristics. *Attention and Performance, 12,* 601–618.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 667–682.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research, 27*(5), 515–539.

Novick, J. M., Kim, A., & Trueswell, J. C. (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. *Journal of Psycholinguistic Research, 32*(1), 57–75.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3,* 410–420.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283–317.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41,* 427–456.

Pollatsek, A., & Rayner, K. (1990). Eye movements and lexical access in reading. In D. A. Balota, G. B. Flores D'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 143–163). Hillsdale, NJ: Erlbaum.

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., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 243–268). New York: Elsevier.

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

Rayner, K., & Sereno, S. C. (1994b). Regression-contingent analyses: A reply to Altmann. *Memory and Cognition, 22,* 291–292.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, R. A., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4,* 21–50.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1188–1200.

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.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147.

Spivey-Knowlton, M., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling effects of referential context and lexical frequency within an integration-competition network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1521–1543.

Sussman, R. S., Campana, E., Tanenhaus, M. K., & Carlson, G. M. (2002). *Verb-based access to instrument roles: Evidence from eye movements.* Poster session presented at the 8th Annual Architectures and Mechanisms of Language Processing Conference, Tenerife, Canary Islands, Spain.

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.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47,* 69–90.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999a). The development of on-line language comprehension abilities in children. *Language, Minds, and Brains: Studies in Languages, 34,* 209–215.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999b). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134.

Notes

1. Some readers will be bothered by the absence of a comma in these sentences. However, similar garden-path effects are found in other structures without "the comma problem."
2. Quasi-first pass time is similar to the regression path duration, except that it does not include time spent in prior regions. It is the sum of all fixations with a region from the time the region is first entered until the reader fixates on a region to its right.
3. Much ink has been spilled on the validity of regression-contingent analyses of reading time (Altmann, Garnham, & Dennis, 1992; Altmann, 1994; Rayner & Sereno, 1994a & b). I remain an agnostic on that issue.
4. None of the semantic anomalies illustrated above are structure-determining. For example, in (6a), syntax dictates that *canary* or *snake* must be the subject of *sings*. Whether or not the animal can plausibly sing is irrelevant to the sentence's structure. In contrast, the sentences in (12), adopted from Boland (1997), illustrate how semantic properties can govern sentence structure in English. A reading event can involve a reader, the thing being read, and a listener. This can be stated in one of two syntactic forms: "The reader read the material to the listener" or "The reader read the listener the material." In (12), the semantic properties of the *wh*-phrase determine whether it will be interpreted as the listener or the material at the verb (recognition of "read" presumably makes its thematic roles available), but the syntactic form of the sentence is not determined until the last word in the sentence. The syntactic position of the *wh*-phrase is represented by the blank line in the examples. Examples (12b) and (12c) both become ungrammatical because *Suzie/Hamlet* satisfies the same set of semantic

constraints as the *wh*-phrase, thereby competing for the same syntactic position. Because semantics determines structure, I would expect to find long reading times for the underlined phrases in (12b) and (12c) compared to the same phrases in (12a) and (12d).

(12) a. *Which poem did the babysitter read Suzie ___?*
b. **Which child did the babysitter read Suzie ___?*
c. **Which poem did the babysitter read ___ "Hamlet"?*
d. *Which child did the babysitter read ___ "Hamlet"?*

In well-known examples like *The spy saw the cop with the revolver* from Rayner et al. (1983), the semantics of *revolver* is not structure-determining in the intended sense. The syntactic analysis of *revolver* itself is unambiguous—it must be the daughter of the PP. Of course, if a PP is constructed and VP-attached at *with*, then *revolver* would be semantically anomalous and the PP might be reanalyzed as modifying *the cop*. However, I am trying to identify the factors determining the initial structure rather than the factors that prompt reanalysis of earlier material.

5. In fact, Griffen and Bock (2000) compared a picture inspection task with a picture description task, using simple line drawings with two participants (e.g., a mouse squirting a turtle). They found an equal number of looks to the two participants during the first 1300 ms of the picture inspection. In contrast, people who were describing the pictures tended to look at the first-mentioned participant during this same time interval. Thus, the nature of the experimental task can clearly influence looking patterns during the initial seconds of picture viewing.
6. Counting co-occurrence frequency in an "anticipatory looking" paradigm presents an interesting problem. The dependent measure is an anticipatory look to a photograph, prior to hearing the target phrase. Thus, the frequency measures that are often used in parsing/reading research are inappropriate. In reading studies, the co-occurrence frequency between a verb and a phrase of a particular class (e.g., a PP beginning with *to* or a phrase that is assigned a particular thematic role) is often used to predict processing difficulty for a phrase of the same class (e.g., McDonald & Shillcock, chapter 5, this volume). In contrast, what we need to know here is, given a particular verb, how likely is the occurrence of an object or person like the one in the target picture. Boland (2003) used two different methods to compute co-occurrence frequency, but it is likely that this research problem will require continued attention.