Eye-tracking and Experimental Syntax

Dave Kush (NTNU) and Brian Dillon (UMass Amherst)

Mini Essay

Eye-tracking-while-reading can be a useful method for investigating the nature of grammatical representations and the commitments made by traditional formal syntactic theories. However, in order to use eye-tracking to this end, experimental syntacticians must have clear hypotheses about how to link grammatical constructs to reading behavior. This requires linking hypotheses between parsing operations and grammatical constructs, and between reading behavior and parsing operations. The ability to articulate these linking hypotheses presupposes a basic understanding of the reading task and how aspects of the task are reflected in eye-tracking measures.

To this end, we provide a rudimentary overview of the psychology of reading and the basics of the eye-tracking-while-reading method. We review a few findings that motivate the assumption that effects of grammatical processing – or at least parsing operations – are reflected in the eye-tracking record. We offer remarks on practical aspects of constructing an eye-tracking experiment, and interpreting its results. Our remarks include cautionary guidelines regarding statistical pitfalls and the tendency to assume too direct a mapping from specific dependent measures to exact parsing operations.

We also discuss issues that face anyone who wishes to align a grammatical model with a parsing model. Parsing behavior (and eye-tracking results) can only be relevant for grammatical theorizing if one assumes some degree of transparency between grammatical constructs and parsing operations. That transparency requires that the analyst explicitly commit to a particular parsing model (or an equivalence class of parsing models) that provides a restrictive inventory of parsing operations to which grammatical constructs can be set into correspondence. This task is complicated by the observation that different parsing models furnish different sets of primitive operations and functions. This entails that predictions about where an analyst might expect to find effects, or what effects constitutes evidence for or against a particular grammatical construct often vary as a function of the specific parsing model assumed.

We offer in-depth case-studies on the processing of A'-dependencies and reflexive anaphors, and discuss how their results reflect on grammatical questions.

1 Introduction

In this chapter we consider some ways that the study of reading behavior using eye-tracking can contribute to experimental syntax. In doing so, we additionally aim to provide a high-level introduction to the technique and review some of the key theoretical background and empirical results in the reading literature. We focus on the eye-tracking-while-reading technique, rather than other techniques that make use of eye-tracking such as the visual world (for an excellent introduction to the use of the visual world paradigm, we refer the reader to Ferreira & Henderson, 2004; for recent examples of its application to syntactic theory, see: Kaiser, Runner, Trueswell & Tanenhaus, 2009; Runner & Head, 2014).

It is important to delimit at the outset what we take 'experimental syntax' to mean. We adopt a relatively traditional view of what syntax, or syntactic theory, encompasses: syntacticians are interested in characterizing the nature of abstract syntactic representations, generally stated at a meta-behavioral level of analysis (e.g., Marr's 1982 *computational level*). That is, we agree with Chomsky (1965; *et seq.*) that syntax is about giving a characterization of human linguistic *competence*.

Traditional syntactic inquiry relies on informal acceptability judgments as the primary method for investigating questions of linguistic competence. Experimental syntax corresponds to the idea that formal experimental design and quantitative analysis of acceptability data can provide an added benefit above and beyond informal methods. Syntacticians may wish to use laboratory techniques as the arbiters of contentious debates that traditional one-off judgment experiments have left unresolved (cf. Phillips, 2010), or they may wish to provide quantitative evidence in support of their empirical claims (see discussion in Gibson & Fedorenko, 2010, 2013; Sprouse & Almeida, 2013; Phillips, 2010). They may also wish to use laboratory methods to provide convergent evidence for an analysis or construct.

Recently grammatically-oriented psycholinguistic research has continued to gain prominence, and the empirical basis of experimental syntax has broadened to include a range of other experimental techniques, including reading measures like eye-tracking. These techniques provide a powerful new tool to investigate the nature of grammatical representations, by investigating how comprehenders compute grammatical relationships incrementally during sentence comprehension. For this reason, researchers interested in using eye-tracking data in the service of experimental syntax must engage questions of linguistic performance. In order to draw theoretically meaningful conclusions from eye-tracking data, it is critical to have a working hypotheses about (i) how the incremental parser operates and how this impacts reading behavior, and (ii) how to link parser operations to grammatical constructs.

Thus while eye-tracking offers - in our view - a technique that promises to shed new light on various issues in syntactic theory, it does impose additional theoretical overhead for the analyst, who must balance commitments both to the nature of the syntactic representations and to the processing mechanisms that assemble those representations in real-time (for similar discussion, see Frazier, 2008; Lewis & Phillips, 2015; Troztke, Bader & Frazier 2013). In what follows we first offer a general – although by no means exhaustive – introduction to the study of eye movements during reading, as well as the eye-tracking-while-reading technique as applied to sentence comprehension. We then turn to a discussion of the theoretical issues that arise when trying to link eye movement behavior to grammatical models. We close by offering in-depth case studies that in our view highlight ways in which eye-tracking data can contribute to syntactic theory.

2 Eye-tracking while reading and syntactic processing

2.1 Eye-tracking: The basic model

A basic working knowledge of the psychology of reading is crucial for understanding the eye-tracking method. As the name suggests, the primary behavior of interest in an eye-tracking while reading experiment is the pattern of eye movements across a span of text during normal reading. In educated and literate adults, reading is an easy and extremely well-practiced behavior. In this respect, eye-tracking offers an ecologically valid counterpart to other psycholinguistic methodologies that rely on visual presentation of linguistic stimuli, such as self-paced reading (Just, Carpenter & Woolley, 1982) or rapid serial visual presentation (Potter, 1984). This feature makes it an attractive methodology for studying language comprehension using visually presented stimuli: the experimental task need not be learned on the spot and should therefore be less susceptible to contamination by 'test-taking' strategies or other task dependent behavior.

Despite the apparent simplicity, the processes necessary to recognize and interpret text during this routine behavior are cognitively complex. Readers must engage visual processing to identify wordforms, link them to appropriate phonological / orthographic codes, access the lexicon and integrate those words into an evolving representation of the sentence and discourse context, all the while continuously engaging in oculomotor programming routines to keep the eyes moving through a text at an appropriate pace (Huey, 1908; Rayner, 1998; Rayner & Pollatsek, 1989; Schotter & Rayner, 2015). Because reading involves coordinating these diverse cognitive processes, the study of reading behavior has engaged a similarly diverse community of cognitive scientists and psychologists (Rayner, 1978, 1998; see Pollatsek & Treiman, 2015, for a broad overview of research in this area). The upshot is a body of research that provides a wealth of empirical data on how reading proceeds for practiced adult readers, but also a firm theoretical model of the reading task itself. This task model in turn provides an important basis for understanding how to link eye movements to higher order linguistic processing. Having such a relatively well-understood task model sets eye-tracking-while-reading apart from other reading methodologies like self-paced reading.

The subjective impression that reading proceeds in a smooth, continuous fashion is misleading. In reality, the eyes move across text by alternating between short periods of relative stability on a single location in the text (*fixations*), and rapid, ballistic movements between those fixations (*saccades*). Fixations generally last between 150 and 500 ms, averaging around 200 to 250ms (Rayner, 1978; Schotter & Rayner, 2015). The saccades between fixations are much shorter, averaging 20-35 milliseconds depending on the length of the movement (Schotter & Rayner, 2015). During a fixation, the visual field can be subdivided into areas of differing visual acuity. The *fovea* is the region with greatest visual acuity, and spans approximately 2° in the center of visual attention. This is followed by the *parafovea*, which has diminished visual acuity and extends to approximately 5° on either side of a fixation; beyond the parafovea is the *periphery* (Rayner, 1998). Normal reading, then, can be thought of as a series of fixations that uptake information about the text as a set of more-or-less static images, each partitioned into distinct regions of visual resolution.

An important theoretical question is what, exactly, determines the pattern of fixations a reader will make across a sentence or text. One critical—but sometimes implicit—hypothesis is that the eyes generally fixate material that is currently in attention; this is referred to as the 'eye-mind hypothesis' (Just & Carpenter, 1984). At present there is a broad consensus that this is more or less correct, though the link between fixation and attention in reading is not strictly one-to-one. Current understanding of the link between attention and fixations in reading is largely built on the work of the late Keith Rayner and colleagues (Clifton, Ferreira, Henderson, Inhoff, Liversedge, Reichle & Schotter, 2017). This link is expressed perhaps most explicitly in the influential E-Z Reader model of reading, a formal model of reading behavior that is built on decades worth of empirical research (Reichle, Pollatsek, Fischer & Rayner, 1998; Rayner, Ashby, Pollatsek & Reichle, 2004; Reichle, Warren & McConnell, 2009). E-Z Reader posits that reading is a serial, attentionally-driven process: a fixation serves the purpose of loading a word in the text into attention, which then feeds lexical access routines as well as higher order syntactic and semantic integration processes. In addition to supporting lexical access, early processing stages (the *familiarity check*) serve the additional purpose of allowing the system to program and execute the next saccade. These claims are not universally endorsed; there exist competitor models that differ in the specific claims they make about this process. For example, the SWIFT reading model rejects EZ Reader's claim that attention is directed serially and sequentially from word to word in favor of an attention gradient that can span multiple words in the text, allowing readers to process several words in parallel (Engbert, Nuthmann, Richter & Kliegl, 2005).

Despite these differences, there is broad consensus that linguistic processing largely determines the duration and pattern of fixations in the text (but cf. McConkie & Yang, 2003). Much research has focused on lexical processing in reading. Research has shown, for example, that fixation durations are reliably longer as a function of a word's frequency (Rayner & Duffy, 1986), its length (Juhasz, White, Liversedge & Rayner, 2008), and its cloze predictability in context (Ehrlich & Rayner, 1981; Staub, 2015); these are all variables thought to influence how readily a word can be accessed in the lexicon. Indeed, the effect of these variables does appear to be due to covert lexical processing, rather than any low-level visual processes; in one striking set of experiments, it was found that frequency effects on fixation duration obtain even when the printed word disappears shortly after a reader fixates it (Liversedge, Rayner, White, Vergilino-Perez, Findlay & Kentridge, 2004; Rayner, Liversedge, Vergilino-Perez & White, 2003). Similarly, readers spend more time fixating ambiguous words than unambiguous words when an ambiguous word has roughly equally available meanings (Duffy, Morris & Rayner, 1988; Rayner & Duffy, 1986). This finding suggests that covert cognitive processes--lexical competition, in this instance--are reflected in longer fixation durations during normal reading.

Overall, there is good empirical support for the view that ease of lexical access is an important determinant of how long readers fixate a word in text. This in turn supports the widely held view that "ongoing lexical access [is] the 'engine' that 'drives' the eyes forward" (Reichle et al., 2009, pp. 4). One notable exception is the landing position for a saccade, which has been argued to reflect purely physical features of the text (but cf. Bicknell, Higgins, Levy & Rayner, 2013).

It is not just lexical processing that drives the pattern of fixations across a text. Difficulty in post-lexical processing, such as syntactic integration or integration into a discourse context, is reflected in eye movements as well (see Boland, 2004; Clifton, Staub, & Rayner, 2008 for helpful reviews). In a seminal study, Frazier and Rayner (1982) showed that reading is significantly disrupted on the underlined regions of the garden-path sentences in (1). First fixation times in the region increased, as did the probability of making a backwards saccade or *regression* in the text.

(1) a. Since Jay always jogs a mile and a half <u>seems like</u> a short distance to him.
b. The lawyers think his second wife will claim the entire inheritance <u>belongs to her</u>.

The authors reasoned that the disruption reflected parser failure and the cost of syntactic reanalysis stemming from syntactic ambiguity. An incremental parser that builds syntactic representations left-to-right and word-by-word faces choice when analyzing the NPs *a mile and a half* and *the entire inheritance* in (1a) and (1b), respectively. Before the underlined regions are read, two analyses are possible: the NPs could either be the direct object of the preceding verbs *jogs* and *claim* (2a), or they could be the subjects of a new clause (2b).

(2) a. [… [$_{VP}$ jogs [$_{NP}$ a mile]] / [… [$_{VP}$ claim [$_{NP}$ the entire inheritance]]
b. [… [$_{VP}$ jogs]] [$_{CP}$ [$_{NP}$ a mile] / [… [$_{VP}$ claim [$_{CP}$ [$_{NP}$ the entire inheritance] …]

According to Frazier's (1978) *Garden Path* theory, the parser is strictly serial, so it must choose a single analysis to commit to. General parsing principles should lead the parser to prefer the direct object analysis in both cases (the principle of *Late Closure in* 1a) and (*Minimal Attachment* in 1b). Subsequently reading the underlined material in both examples disambiguates the parse to the dispreferred separate-clause analysis, which necessitates costly revision and reanalysis. These results provide evidence that the parser incrementally makes syntactic commitments that correspond to representational distinctions in our formal theory and that the process of making and managing these commitments is reflected in the eye movement record.

Since this seminal work, there have been well over a hundred studies that demonstrate that 'garden-pathing' of this sort reliably impedes a reader's progress through the text in eye-tracking-while-reading experiments (Clifton et al., 2008). Moreover, it is not only the resolution of a syntactic ambiguity of this sort that can create an increase in fixation durations and regressive eye movements. The recognition of syntactic and semantic anomalies has a similarly deleterious effect on reading. For example, Pearlmutter, Garnsey & Bock (1999) and Deutsch &

Bentin (2001) provided early demonstrations that fixation durations, as well as the probability of making a backwards regression, increased when readers encountered a number agreement error on the verb in English, or a gender agreement error in Hebrew (respectively). In a similar vein, Rayner, Warren, Juhasz & Liversedge (2004) showed that the various types of implausibility in (3) can impede reading:

(3) a. The man used a strainer to drain the thin <u>spaghetti</u> …
b. The man used a blow-dryer to dry the thin <u>spaghetti</u> …
c. The man used a photo to blackmail the thin <u>spaghetti</u> …

(3b) and (3c) present two different types of implausibility: (3b) is implausible in light of our world knowledge, whereas (3c) is impossible and arguably reflects a selectional restriction that *blackmail* places on its object (i.e. that it must be animate). Interestingly, these different types of implausibility impacted fixations on the critical word differentially: the anomaly in (3c) was registered from the very first fixation on *spaghetti*, unlike in (3b), and readers regressed more frequently in (3c) than (3b). To the extent that these two types of implausibility play out differently in the eye movement record, there is a plausible argument to be made that these anomalies have different cognitive statuses.

Early versions of E-Z Reader focused primarily on modeling the role of lexical access in reading, but more recent versions have integrated explicit, post-lexical processing stages into the model in order to capture effects of syntactic or semantic integration like those described above (Reichle, Warren, & McConnell, 2009). Although this model remains fairly agnostic about the nature of the parsing and interpretation events that occur in this post-lexical processing stage, one strand of current research seeks to build more explicit links between formal models of parsing and models of eye movements (e.g., Engelmann, Vasishth, Engbert & Kliegl, 2013; Vasishth, Bruessow, Lewis & Drenhaus, 2008; Vasishth, von Der Malsburg & Engelmann, 2012). In our view, further research on the link between parsing models and eye movements is of critical importance for the use of eye movement measures in experimental syntax. As we discuss below, the degree to which eye-tracking methods will prove informative for syntactic theory depends on making explicit the links between the grammatical model and the parsing model, and the parsing model and the eye movement model. Despite the widespread recognition that syntactic and semantic processes are reflected in the eye movement record, we believe that much work remains to be done on linking parsing models and eye-tracking models.

2.2 Eye-tracking: The practice

We now turn to practical questions of how an eye-tracking-while-reading experiment proceeds, and how the data it produces are analyzed. In a canonical eye-tracking-while-reading experiment, a participant is presented with a single stimulus sentence on the screen, usually in a monospace font such as Courier or Monaco. Typically, the entire sentence will be presented on screen, and the participant will read at their leisure until they are satisfied that they have understood the sentence. At that point, they will proceed to a comprehension question or the next sentence by pressing a button on a response pad. Variations of this basic setup involve presenting multiple lines of text, offering the participant additional context for the stimulus sentence, or perhaps an additional secondary task above and beyond simple reading for comprehension (Clifton, 2013).

During this passive reading, an eye-tracker is used to monitor the duration and location of fixations and saccades across the text. The raw data from an experiment essentially consist of a record of the duration and location of the fixations that a participant makes on a given sentence. Analysis of these data typically involves i) specifying a *region of interest* (ROI), or a portion of the sentence within which the fixations will be analyzed and ii) calculating a number of derived dependent measures that summarize the pattern of fixations within a given span of text. Suppose an analyst was interested in measuring the disruption caused by subject-verb agreement failure in the ungrammatical sentence (4):

(4) The lonely hiker on the mountain were getting tired and decided to take a break.

Under the assumption that effects of interest would at least be localized to the region where agreement failure becomes apparent, the analyst might choose the ungrammatical auxiliary *were* as a region of interest, potentially including the following word as part of the ROI. Since effects are not always immediately localized to a critical region, one might also choose to analyze a *spillover* region following the critical region (*tired* in this example).

Having specified a region of analysis (or several), various dependent measures are computed for each ROI. The *first fixation duration* refers to the duration, typically expressed in milliseconds, of the first fixation that lands in the ROI. *First pass reading time* is the sum of all fixations on an ROI before the reader exits it to the left or to the right; this is sometimes called *gaze duration* when the ROI comprises only a single word. *First-pass regressions out* is the proportion of trials where readers make a backwards saccade or regression from the ROI after their first pass. *Go-past time* or *regression path* is the sum of all fixations that occur from the first fixation in an ROI until the reader exits it to the right; this measure includes fixations made on previous stretches of the text after a regression. Together, these measures comprise so-called 'early' measures of reading behavior (though see below for caveats concerning a tidy early/late dichotomy), as they jointly index all processing that occurs up until the reader moves to the right of a given region. Later measures include *re-reading time*, which is the sum of all fixations in an ROI after the reader moves past it for the first time, and *total time*, which is the sum total of all fixations on a given ROI, regardless of when they occur during the course of reading a sentence. These derived dependent measures are some of the most common reported in eye-tracking studies. However, eye-tracking data analysis remains an active area of investigation. Some more recent, but less commonly used measures include scanpath analysis (von der Malsburg & Vasishth, 2011, 2013), and cumulative progression analysis (Scheepers, Hemforth, Konieczny & van Gompel, m.s.; cited in Kreiner, Garrod & Sturt, 2008). These measures have strengths and weaknesses that complement more traditional analysis. Interested readers should consult these references for further information.

As the preceding makes clear, the eye-tracking analyst faces numerous substantive analytic choices. This presents a somewhat perilous situation: in the absence of clear theoretical guidance on the choice of a region of analysis, or the appropriate dependent measure, there are a number of delicate statistical issues. First, the multiplicity of dependent measures typically analyzed in an eye-tracking experiment evidently leads to a potentially serious multiple comparisons problem; eye-tracking researchers typically declare an 'effect' after observing a significant effect of a manipulation in a single dependent measure (von der Malsburg & Angele, 2017). This can dramatically increase the probability of a Type I error; a simple heuristic to avoid this is to ensure that any critical effect is found in more than one dependent measure, but other measures to correct for multiple comparisons such as the Dunn-Bonferroni correction may also be applied (von der Malsburg & Angele, 2017). Second, we believe there is no clear theoretical guidance (yet) on the optimal choice of analysis regions. This leads to another statistical problem: if researchers choose their regions of analysis after the fact, they may inflate their experiment-wise Type I error rate (i.e. they may unwittingly enter 'the Garden of Forking Paths'; Gelman & Loken, 2013). In our view, the situation at present is one where choice of analysis region should be viewed as exploratory. There is substantial study-to-study variation in where and when a given parsing effect will arise (Clifton et al., 2008), and so it is not (yet) possible to predict a priori exactly where a given effect will arise. In light of this, it seems unavoidable that researchers will sometimes need to examine multiple distinct regions of analysis for a given experiment. In this case, it is important to present fully and transparently the results of such exploration in published work so that the distinction between confirmatory and exploratory analyses is clear; Pearlmutter et al. (1999) provide an excellent example of this approach.

2.3 *The time course of grammatical processing*

Perhaps one of the most oft-cited benefits of the eye-tracking-while-reading methodology when applied to sentence processing is its temporal resolution: the analyst can examine the time course of processing from the first fixation on a critical region all the way up to any re-reading of that region. Given the relative ordering, it is customary to refer to first fixation and first pass as 'early measures', and second pass / re-reading and total time as 'late measures' (Rayner, Sereno, Morris, Schmauder & Clifton, 1989). Go-past and regressions out are sometimes labeled early, sometimes late (Clifton et al., 2008). This rough distinction allows some insight into the processes of interest: to the extent that a given manipulation is reflected in first fixation or first-pass measures, it seems safe to say that it reflects the initial processing of the region of interest. On the other hand, effects that only occur in late measures like re-reading are less likely to result from early processing of the target region.

However, it is important to resist reifying the early-late distinction. As mentioned above, there are some measures like go-past times whose classification is up for debate. More importantly, however, there seems to be a many-to-one relationship between parsing processes and eye-tracking measures. This was clearly demonstrated by Clifton, Staub and Rayner (2007), in a review of 100 peer-reviewed papers that used eye-tracking to investigate 'higher-order' processes, such as resolving a syntactic ambiguity, recognizing a syntactic or semantic anomaly, or processing syntactic or referential complexity. The survey indicated that for any given manipulation or type of anomaly, there was substantial variation in where, and on what measure, it impacted eye movements. For instance, the authors observed that the processing of a syntactic anomaly impacted first pass times and regressions out on both the critical region and on the spillover region, as well as late measures like total time and re-reading. The subsequent literature has been similarly varied: for example, the recognition of an agreement error has been shown to impact go-past times at the critical region and the spillover (e.g. Kreiner, Garrod & Sturt, 2013), first pass and total reading times (Dillon, Mishler, Sloggett & Phillips, 2013), and regressions out (Dillon et al., 2013).

Despite this uncertainty, it is clear that grammatical processing does impact eye movements, and that it can do so quite rapidly. One of the most common experimental paradigms for studying grammatical processing in comprehension is the so-called violation paradigm; in other well-known measures for studying comprehension, such as the event-related potentials (ERP) technique or self-paced reading, this paradigm is widely used. It has been less common in eye-tracking-while-reading. In the violation paradigm, the experimenter will introduce a syntactic (or semantic) anomaly at some position in the sentence, and measure when, and to what degree, the recognition of the anomaly impacts reading comprehension; Rayner et al. (2004)'s study on the various types of implausibility offers one such example. Another example comes from Kreiner, Garrod & Sturt (2013), who looked at examples such as (5):

(5) a. The family / families definitely and undeniably <u>wish</u> <u>to avoid</u> a court trial.
b. The widow / widows definitely and undeniably <u>wish</u> <u>to avoid</u> a court trial.

Kreiner and colleagues introduced an agreement error on the critical verb *wish* by manipulating the number marking on the subject head noun; the following two words served as a spillover region. Kreiner et al. wanted to know whether notionally plural collective nouns like *family* would mitigate any processing difficulty created by a mismatched verb. Interestingly, they observed that both collective nouns like family and non-collective nouns like *widow* imposed a similar penalty on a mismatching verb in early measures; in particular, they both created a slow-down of approximately 80ms in go-past times at the verb. Thus, in an arguably 'early' measure (go-past times), a mismatched verb had an impact on the eye movement recording, slowing fixation durations and/or triggering additional regressive saccades and re-reading.

In itself, it is difficult to characterize a mismatch penalty in go-past times as extrinsically 'early' in processing for all the reasons detailed above: there is just too much variation in where comparable mismatch effects are observed, study to study, to draw firm conclusions of this sort. However, within an experiment, it is possible to distinguish the relative time course of the grammatical processing for a given comparison. Indeed, this is exactly what Kreiner and colleagues did. For those same stimuli, the mismatch penalty they observed looked importantly different in second pass reading times. In particular, there was no longer any reliable mismatch penalty for collective nouns, while the penalty for non-collective nouns persisted (and indeed appeared numerically greater than in go-past times). Thus, within this experiment, there was a reliable difference in the time course of the mismatch penalty when comparing the different noun types: non-collective nouns imposed an early and long-lasting mismatch penalty; collective nouns imposed an early penalty that dissipated in later measures. Like Rayner et al. (2004), these data point to a distinction between notionally plural collective nouns and non-collective nouns with respect to how they enter into agreement dependencies with the verb. Of course, the interpretation of such a distinction is up for debate: Kreiner and colleagues suggest that the semantic information is available only at a delay, and so can come online to resolve an agreement mismatch only in later processing stages. What is clear, however, is that eye-tracking affords a comprehensive picture of the comprehension process as it unfolds, which allows the analysis to temporally dissociate plausibly distinct subcomponents of the parsing process. To the extent that parsing processes make direct contact with the representational vocabulary afforded by the grammar, such dissociations have the promise to bear on questions of experimental syntax.

Though the precise ways in which higher-order grammatical processes are reflected in the eye-movement record remain poorly understood, the same is not true of lexical processing. The lexical factors that control eye movements and the measures that they impact are comparatively well understood. Thus, eye-tracking may be particularly useful for experimental syntacticians whose theoretical questions concern the dividing line between the lexical and grammatical processes. Explicit comparisons of how lexical ambiguity and syntactic ambiguity play out in eye-tracking-measures reveal some similarities, and some differences; the differences have been used to argue against processing models that rely on lexical storage and retrieval of syntactic frames to assemble syntactic structure in comprehension (Traxler, van Gompel & Pickering, 1998). For example, we noted above that comprehenders spend more time fixating ambiguous words than unambiguous words when the ambiguity is relatively balanced (Duffy, Morris & Rayner, 1988; Rayner & Duffy, 1986). The same cannot be said of syntactic ambiguity: processing syntactic ambiguity does not reliably cause reading time slowdowns. For example, the verb *found* can take either an NP or CP complement, as shown in (6a, 6b), but there does not seem to be reliable evidence that the availability of two parses slows participants down at the verb (see Hare, McRae & Elman, 2003 for the original experiment and subsequent discussion in Clifton et al., 2007; Clifton & Staub, 2008).

(6) a. The psychology students found the book in the bookstore.
    b. The psychology students found the book was written poorly.

While evidence for costs of ambiguity is scarce, there is good evidence that syntactic ambiguity often conveys a processing *advantage* in eye-tracking measures, speeding reading times relative to unambiguous sentences. This effect is known as the *ambiguity advantage effect* (Traxler, Clifton & Pickering, 1998; van Gompel, Pickering & Traxler, 2000, 2001; van Gompel, Pickering, Pearson & Liversedge, 2005). The observation that lexical ambiguity slows reading times, while syntactic ambiguity speeds reading times, has provided an important challenge for syntactic processing models that rely on lexical storage and retrieval to assemble syntactic structures in incremental sentence processing (Clifton & Staub, 2008; Traxler et al., 1998; van Gompel, 2000; but see Green & Mitchell, 2006; Vosse & Kempen, 2009 for counterarguments).

3 Aligning Grammar and Parser

Processing models and results such as those discussed in the previous section are descriptions expressed at the algorithmic level: they concern how readers create, maintain, and update incremental syntactic representations over the course of processing a sentence. By contrast, syntactic theories are generally formulated at a higher level of abstraction (Sprouse & Lau, 2013). This gap needs to be bridged in order for eye-tracking data to be put to service for syntactic theorizing. In order for eye-tracking results to have bearing on the form of our syntactic theory, we need linking hypotheses that map (i) grammatical constructs to parsing operations and (ii) parsing operations to dependent measures like the various reading time measures discussed above. The issue of how to link grammatical representations with parsing operations is a very difficult issue: we highlight some of the basic issues we see (we point the reader to Boland 2005 for additional extended discussion, and Marantz, 2005 for a somewhat more optimistic take on the enterprise).

Being able to draw inferences about the grammar from real-time results presupposes the intuitive conjecture that there exists a (non-trivial) correspondence between operations and constructs of the grammar and those of the parser. Complex grammatical representations, for example, should require the parser to conduct more complex computations all else equal, while less complex representations should require less computation. More complex computations should require greater effort, which should in turn manifest as an increase in reading times (or some other similar measure).

The degree of correspondence between the grammar and the parser is to some extent an open empirical and theoretical issue and the extent of correspondence that the analyst assumes has consequences for how to interpret parsing data. The strongest assumption to make is that there is a maximally transparent or *isomorphic* relation between grammatical and parsing operations. The now disfavored Derivational Theory of Complexity (Chomsky & Miller, 1963) assumed such strong isomorphism (see Fodor, Bever & Garret, 1974 for early arguments against the DTC; and Phillips, 1996 for a different perspective). Weaker variants of the transparency thesis give up on strict isomorphism in favor of some kind of weaker homomorphism, which permit many-to-one mappings between grammatical and parsing operations in either direction.

One well-known example of homomorphism is Berwick and Weinberg's (1984) *weak type transparency* condition, which holds that grammatical rules and structures are transparently instantiated in the parser, but the parser may make additional distinctions above and beyond those imposed by the rules and representations in the grammar. We suspect that most theorists nowadays endorse homomorphism at least implicitly, but the exact details vary (see Marantz 2005, Lewis & Phillips, 2013, 2015).[1]

If one has a formal model that clearly delimits both the basic inventory of parsing operations and the primary determinants of processing costs, one can align grammatical constructs to parsing operations (either one-to-one as in the case of isomorphic mappings, or one-to-many in the case of homomorphism). In turn, one can make relatively straightforward predictions about when and where effects of a grammatical manipulation should emerge in the eye-tracking record.

It is important to note, however, that such *predictions are often only valid with respect to the particular parsing model assumed*. Different models of the parser often apportion the costs of complexity differently across various computationally divergent operations and therefore make different predictions about where and when grammatical effects are supposed to emerge. Therefore, it is important for the experimental syntactician to commit to a specific model (or equivalence class of models) from the outset.

The problem that the syntactician will encounter here is that there are many models on the market, many of which vary quite substantially in their details and parameters, and consequently the explanations that they support. Consider a few choice-points that one is faced with when choosing a parsing model: One must decide, whether the parser is *serial* (i.e., it only pursues a single syntactic analysis for an input sentence at a given time), or *parallel* (it can consider multiple possible parses simultaneously). One must also be clear about the memory architecture that subserves the parser: How does it encode and store information? What are its capacity limits? How is information maintained and accessed? Further, one has to determine how *predictive* the parser is: can it build syntactic structure or compute dependencies before it has unambiguous bottom-up evidence? Each of these choices have theoretical and practical consequences for interpreting results and making predictions about incremental processing, since they determine, in part, the inventory of parsing operations that grammatical constructs can be set in correspondence to, and because they each come with their own grammar-independent complications that must controlled for. For example, as pointed out by Phillips & Wagers (2007), essentially all of the psycholinguistic evidence adduced against the existence of traces/gaps (e.g.

---

[1] We should note that not all researchers agree that transparency is warranted: Townsend & Bever (2001), for example, propose that initial parsing is carried out using a suite of 'quick and dirty' heuristics that are only indirectly related to full-fledged grammatical representations which are computed at some subsequent point. Other researchers working within the *Good Enough Parsing* (GEP) framework have gone so far as to argue that the end state of parsing needn't always be a well-formed syntactic representation (see Ferreira & Patson, 2007). It is clear that adopting either of these hypotheses makes direct inference from parsing measures to grammatical conclusions incredibly difficult. For the purposes of this chapter we assume that there is a degree of transparency (see Lewis & Phillips, 2015; Phillips, 2013 for further discussion).

Pickering & Barry, 1991) rests on the assumption that parsing is bottom-up. The findings are equally consistent with a parser that *predictively* posited the existence of a trace before its linear position (see also Aoshima, Phillips & Weinberg, 2004; Gibson & Hickok, 1993; Gorrell, 1993; Crocker, 1994).

Given that interpretation so often depends on model assumptions, we advise aspiring experimental syntacticians to adopt independently established and explicit parsing frameworks whenever possible. This is for two reasons: First, explicit parsing frameworks allow clearer mapping hypotheses between parser and grammar, which can help in refining predictions and testing more precise theoretical hypotheses. Second, results are more easily falsifiable and evaluable. As we discuss later, we have couched much of our own work within the general cue-based parsing framework (e.g., Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006; McElree, 2006) partially for these reasons. The cue-based framework posits a serial left-corner parser with an extremely limited active memory. The parser's extremely limited focus of attention (a capacity of one or two 'chunks' at most) entails that the parser must frequently *retrieve* previously-processed chunks from a separate content-addressable memory store based on their features (associative cues). Parsing costs within this framework are largely tied to these retrieval procedures and questions of grammatical representation are often formulated in terms of encoding: how is grammatical information encoded in features that can be used as retrieval cues? Theorists working within this framework have leveraged this perspective to ask sophisticated questions about the precise nature of syntactic representations constructed in incremental processing: for example, Arnett and Wagers (2017) use this perspective to investigate what features define a subject phrase in English.

Obviously, the biggest drawback to formulating hypotheses relative to highly precise models is that not everyone shares the same assumptions. Results that depend too heavily on a controversial assumption may be dismissed by theorists working within other frameworks. Hopefully, as parsing models are elaborated and consensus is reached on at least some parameters of the parser, this problem should recede. The success of experimental syntax therefore depends in part on finding answers to questions about the parser that might seem orthogonal to grammatical questions in and of themselves.

4. Case Studies

In the remainder of this chapter we consider two areas in syntactic theory where eye-tracking results may inform debates in syntactic theory: (i) the origin of island constraints and (ii) constraints on the interpretation of reflexive pronouns. We spend some time outlining the logic of the studies for two reasons. First, we wish to familiarize readers with some independent and well-established effects commonly used in eye-tracking studies and explain how such effects can be used to probe representational commitments at the level of syntax.[2] Second, we also wish to

[2] There exists another approach to using eye-tracking measures to probe syntactic knowledge that proposes to draw inferences about syntactic representation by reading effects of

highlight the (sometimes complicated and tenuous) chain of inference required to connect empirical results to theoretical claims.

4.1 Islands

Certain syntactic domains are *islands* for long-distance dependency formation (Ross, 1967). For example, it is possible in many languages to relate a *wh*-phrase like *what* to a gap inside a declarative complement clause (7a), but not inside a relative clause (7b), or a subject (7c).

(7) a. What did Knut say [$_{CP}$ that the kid read __] ?
b. *What did Knut know the kid [$_{RC}$ that read __] ?
c. *What did Knut say that [$_{SUBJ}$ the kid reading __] was amused ?

Why should some syntactic domains, but not others, block the extraction of *wh*-phrases? One influential line of thought attributes island effects to innate syntactic constraints (e.g. Ross, 1967; Chomsky, 1972). Other researchers reject a syntactic account, positing instead that islands are reducible to limitations of extra-linguistic cognitive domains like memory (Deane 1991; Kluender 1993, a.o.). Recent studies have used experimental methods to argue for one position or the other, often using reading measures like self-paced reading (for recent summaries we refer the reader to Phillips, 2013 and Wagers, 2013). However, relatively few have used eye-tracking to study how islands constrain incremental filler-gap processing. We focus on two such studies.

In order to interpret a displaced phrase like a *wh*-word (henceforth, a *filler*) readers must connect the filler to its base-position (its *gap*) that falls somewhere later in the sentence. One challenge that readers face is that the location of the gap is often uncertain when the filler is first encountered. A filler can, in principle, be linked to a number of different plausible positions, as in (8).

(8) Knut asked *who* …

a.	___ liked brown cheese.	[SUBJECT GAP]
b.	Marit saw ___.	[DO GAP]
c.	Torgunn gave the cheese to ___ .	[IO GAP]

(representational/operational) complexity directly off of reading time measures. For example, Boston and colleagues (Boston et al. 2008) showed a significant correlation between reading times and grammatical complexity (as quantified using a surprisal metric, Hale, 2001). Such correlations between some experimental measure and complexity can be used to argue for the utility of syntactic representations in models of processing, and it might also be seen as a way to arbitrate between different formalisms. Other authors in this volume (Brennan, *neurophys chapter*) use a variant of this correlational approach to argue for neural correlates of syntactic processing. While we think that such work is an important first step for neurophysiological work, we believe that it rarely permits us to ask the fine-grained questions that most syntacticians are concerned with: most competing theories agree on the general loci of complexity, but differ in the details (consider the fact that Boston and colleagues, 2008 found that both dependency and phrase-structure grammars were good predictors of reading times).

Psycholinguistic research suggests that readers manage this incremental uncertainty by predicting gap positions. According to this *active filler strategy*, readers pre-emptively posit a gap as soon as possible, even before receiving bottom-up confirmation of the true gap site (e.g. Stowe 1986; Frazier & Clifton 1989).

Early eye-tracking evidence for active filling comes from Traxler and Pickering (1996)'s Experiment 2, where English participants read sentences like (9). Sentences contained RCs whose head (*book*/*city*) was ultimately linked to an oblique gap (__). Sentences also contained an optionally-transitive verb (*wrote*) intervening between the filler and the true gap. According to the active filling hypothesis, participants should initially interpret the filler as the object of *write*. Traxler and Pickering used a *plausibility mismatch manipulation* to determine whether participants did so. They manipulated whether the filler was a plausible object of the intervening verb (one can write a *book*, but not a *city*). They reasoned that if readers posit a gap position after *write,* then readers should experience difficulty if the filler was an implausible object but not to a plausible one.

(9) We like the **{book / city}** that the author *wrote* unceasingly and with great dedication about ___ …

The researchers observed a *plausibility mismatch effect* at *wrote* in (9): Gaze durations were longer on the verb when the filler was *city* than when it was *book*. Based on this effect, Traxler and Pickering concluded that comprehenders were active gap-fillers (corroborating earlier findings, e.g. Stowe, 1986).

The experiment also contained two additional conditions relevant to islands. The conditions featured the same plausibility-mismatch manipulation and the same verb intervening between the filler and gap. The only difference was that the intervening verb was embedded in an island: the embedded subject NP (10). The researchers wanted to know whether readers would blindly associate an un-integrated filler with *any* intervening verb, or whether readers would only posit gaps in positions that were potentially grammatical.

(10) We like the **{book / city}** that [$_{\text{SUBJECT}}$ the author who wrote unceasingly and with great dedication] saw ___ …

Traxler and Pickering found no plausibility-mismatch effect at *wrote* in (10), which indicated that participants suspended active filling inside the subject. The authors took this as evidence that islands constrained parsing behavior.

As pointed out by Phillips (2006), Traxler and Pickering's (1996) results are equally compatible with two different explanations: active filling might be suppressed inside islands because A'-dependencies into islands are grammatically excluded, or because active-filling inside islands taxes memory resources too greatly. If the first explanation were correct, we would have (indirect) evidence that islands are grammatical in origin. But how might we tease the two possibilities apart?

Reductionist accounts predict a blanket ban on active filling in domains coarsely categorized as islands (subjects, RCs, etc.). The ban is expected to hold cross-linguistically, under the assumption that humans possess essentially the same working memory resources, irrespective of their native language. In contrast, the grammatical approach allows for some wiggle room: if there exist languages that allow grammatical dependencies into apparent islands, we should expect active filling in exactly those domains. We consider one experiment that we think provides weak support for the grammatical account. (We also point the reader to Phillips 2006, where a different version of the same argument is made in English.)

Mainland Scandinavian languages like Swedish appear to allow some long-distance dependencies into RCs (Maling & Zaenen 1982; Engdahl 1996, a.o.). If RCs are not islands in Swedish[3], then, according to the logic above, Swedish readers should pursue active filling inside of RCs, all else equal. Tutunjian, Heinat, Klingvall and Wiklund (2017) tested this hypothesis in Swedish using a modified version of Traxler & Pickering's (1996) design. They tested whether readers temporarily interpreted a topicalized NP filler (*såna där möbler/flyttlådor* in 11, 12) as the object of an optionally-transitive verb embedded inside an RC (*renoverade*). The verb of interest preceded the filler's true gap site (the object position of *bära* 'to carry'). Tutunjian and colleagues manipulated whether the filler was a plausible object of the RC-internal verb or not: *möbler* ('furniture') can be 'renovated' in Swedish, but *flyttlådor* ('moving boxes') cannot. They also manipulated the position of the critical RC. In RC-Object conditions (11), the NP containing the RC was the object of the matrix verb *bad*[4] ('asked'), while in RC-Subject conditions like (12) the NP was the matrix subject.

(11) RC-Object

Såna där **{möbler /flyttlådor}** bad jag [$_{NP}$ en kollega [$_{RC}$ som *renoverade* (_)
Such there furniture/boxes asked I a colleague who renovated
på land-et]] att bära __ efter match-en i söndags.
on land-def to carry after match-def last Sunday
'Such furniture, I asked a colleague that renovated (_) in the country to carry __ after the match on Sunday.'

(12) RC-Subject

[3] The islandhood of RCs in Mainland Scandinavian languages is a point of ongoing debate. Some advocate a position that RCs are, across-the-board, non-islands (e.g. Allwood 1982, Engdahl 1996), while others have argued that a more fine-grained analysis is required to distinguish between acceptable and unacceptable RC-spanning dependencies (Kush et al. 2017). For the purposes of this section we adopt the view that they are non-islands in order to illustrate the point.

[4] The matrix verb *bad* immediately follows the topicalized object because Swedish is a V2 language (Holmberg & Platzack, 1995).

Såna där **{möbler /flyttlådor}** bad [$_{DP}$ en kollega [$_{RC}$ som *renoverade* (__)
Such there furniture/boxes asked a colleague who renovated
på land-et]] mig att bära __ efter match-en i söndags.
on land-def me to carry after match-def last Sunday
'Such furniture, a colleague that renovated (_) in the country asked me to carry __ after the match on Sunday.'

The researchers reasoned that if Swedish RCs are not islands, participants should actively interpret the filler as the object of *renoverade* in (11), yielding a plausibility mismatch effect. They further reasoned that active filling should not occur in (12) because subjects are strong islands in Swedish as in English (see Engdahl 1982). Thus, the RC-subject conditions were intended to serve as controls where we would expect no mismatch effect.

The results of the experiment were somewhat complex: Gaze duration and total reading times were longer on *renoverade* when the filler was implausible than when it was plausible. Although the mismatch effect was numerically larger in (11) than in (12) for both measures, the plausibility × RC-position interaction was not significant. The authors found one effect that they argued supported a distinction between RC-Object and RC-Subject conditions: a three-way plausibility × RC position × trial order interaction in gaze duration. This interaction indicated that there was a large plausibility mismatch in the RC-Object conditions on early trials, but this effect was extinguished over the course of the experiment. In RC-Subject conditions, there was no early effect of plausibility, but suggestions of one began to emerge near the end of the experiment.[5]

Overall, the experiment provides suggestive evidence that active gap-filling is not suspended inside Swedish RCs. The results minimally show that active gap-filling inside RCs (and subjects) is not precluded due to resource limitations. The results might also be interpreted to support the non-islandhood of Swedish RCs in general. If RCs are not islands in Swedish, then the results might support the hypothesis that grammatical acceptability controls active gap-filling. Moreover, if island effects do not reflect resource limitations and can vary cross-linguistically, then we the results constitute an indirect argument that islands require a grammatical explanation (though perhaps not one tied to inviolable universal constraints). We point out, however, that these conclusions should be treated with care because aspects of the results – particularly the suggestion of active gap-filling inside subjects – weaken the strength of the findings.

4.2 Reflexive Processing

---

[5] We suspect it is possible that active gap-filling may be possible in Swedish subjects with finite RCs due to the rather permissive parasitic-gapping possibilities in the language (see Engdahl, 1983).

In an influential study, Sturt (2003) investigated the real-time resolution of English reflexive anaphors such as *himself*, with the goal of determining at what stage of incremental processing comprehenders apply Binding Principle A. Sturt sketched two general options: Knowledge of Principle A could apply as an *initial constraint* on the consideration of potential antecedents, restricting the set of possible antecedents to those licensed by the grammar; results from cross-modal priming seemed to support this view (Nicol & Swinney, 1989). Alternatively, Principle A could apply as a *late filter* on antecedent consideration. Under this second option, the parser might initially consider (feature-matching) NPs in grammatically inappropriate positions as potential antecedents, but then rescind this consideration at a later point in processing; some results from self-paced reading seemed to support this view (Badecker & Straub, 2002).

Sturt proposed that eye-tracking could help differentiate between these two hypotheses. Sturt employed a *gender-mismatch* paradigm in an eye-tracking-while-reading study. The logic of this paradigm is as follows: if a reader encounters a reflexive that lacks a feature-matching antecedent in the local context, the parser will fail or otherwise experience integration failure (e.g. Reichle et al., 2009; Vasishth et al., 2008). Thus, reflexives without a feature-matched antecedent will be read more slowly than anaphors that have an antecedent, owing to the increased likelihood of integration failure (although other linking hypotheses between the parser and reading times in this paradigm are possible: see Jaeger, Engelmann & Vasishth, 2017; Nicenboim, Engelmann, Suckow & Vasishth, m.s.; Patil, Lewis & Vasishth, 2016). Thus, we expect longer reading times at the reflexive *himself* in (7) than we would expect to find on *herself* in the same position.

(7) The girl hurt *herself/himself* wrangling the reindeer.

(Gender)-mismatch effects of this type can be used to probe whether certain NPs are 'visible' to the parser as potential antecedents or licensors. Sturt's (2003) Experiment 1 manipulated the gender-match between a reflexive (*himself/herself* below) and two c-commanding NPs. The subject of the reflexive-containing embedded clause (*the surgeon*) was *accessible* because it was local enough to bind the reflexive according to Principle A. The reflexive either matched the stereotypical gender of the accessible NP (*himself*) or mismatched it (*herself*)[6]. The second NP was a pronoun in the matrix subject position of the second sentence. The pronoun was coreferential with a name introduced in the first sentence (*Jonathan/Jane*) and either matched or mismatched the reflexive in stereotypical or definitional gender. As this higher NP was not local to the reflexive (not contained within its immediate finite clause), it was ruled out as a potential antecedent by Principle A. Sturt reasoned that if the parser considered grammatically inappropriate antecedents, the reflexive in (8c) should be processed differently

[6] Sturt (2003) used stereotypical gender in order to avoid presenting people with fully ungrammatical sentences in an eye-tracking experiment. In eye-tracking paradigms, stereotypical gender violations and definitional gender violations lead to similar gender mismatch effects in early measures (Kreiner, Garrod & Sturt, 2008). In later studies (e.g. Dillon et al., 2013; Parker & Phillips, 2017), definitional gender violations or number violations were also used.

than in (8d). Gender-match with the inaccessible NP might facilitate or inhibit processing of the otherwise unlicensed reflexive relative to the condition where the inaccessible NP did not match.

(8)

(a) Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked himself with a used syringe needle.
(b) Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked himself with a used syringe needle.
(c) Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked herself with a used syringe needle.
(d) Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked herself with a used syringe needle.

Sturt found that first-fixation, first-pass, and regression-path times were longer for a reflexive that mismatched the accessible NP (8c,d) than for one that matched (8a,b). Gender-match between the reflexive and the inaccessible NP did not reliably influence these measures at the reflexive. However, the gender of the inaccessible NP did appear to affect later processing: second-pass reading times at the reflexive were shorter when both NPs matched the reflexive (8a) and longer when only the inaccessible NP matched (8d). In later regions, the presence of a matching inaccessible NP facilitated second-pass times, suggesting that the non-local, c-commanding NPs may be considered at some stage of processing.

In a second experiment Sturt (2003) again used a gender-mismatch design to test the effect of an inaccessible NP in a different structural position. In Experiment 2 the inaccessible NP was embedded inside a relative clause attached to the accessible NP. In this position the NP was not a grammatically acceptable antecedent for the reflexive because it did not c-command the reflexive.

(9) The surgeon who treated Jennifer/Jonathan had pricked himself/herself with a used syringe needle.

As in Experiment 1, participants' first fixation and first-pass times were longer in the critical region when the reflexive did not match the stereotypical gender of the accessible NP. Unlike Experiment 1, the gender of the inaccessible NP had no observable effect on any early or late measures.

From these findings, Sturt concluded that Principle A was deployed as a *defeasible* filter: it constrained antecedent selection in early processing, but could be overridden in subsequent processing stages if, for example, countervailing discourse constraints made the inaccessible antecedent particularly tempting. Since Sturt's seminal finding, there has been an intense interest in the processing of reflexives, and in particular, intense interest in the question of whether Principle A is applied as a 'hard' constraint on antecedent retrieval (Dillon et al., 2013), or if instead, Principle A provides only one constraint among many used to select an antecedent (Badecker & Straub, 2002; Jaeger et al., 2017; Patil et al., 2016; Parker & Phillips, 2017; Sloggett, 2017). At present, this debate between these two views continues; see reviews in Dillon (2014), Sturt (2013), and especially Jaeger et al. (2017). A recent meta-analysis by Jaeger et al (2017) suggests that there is very little evidence for interference from inaccessible antecedents on reflexive dependencies; at the same time, recent work by Parker and Phillips (2017) and Sloggett (2017) calls this conclusion into question. We return to these results below.

While recent evidence calls into question the strong conclusion that Binding Theory is rigidly deployed as a 'hard' constraint on antecedent access (Dillon et al., 2013; Dillon, 2014), this body of literature does provide evidence that knowledge of Principle A constrains the parser's earliest attempts to resolve a reflexive dependency. Given this, it is interesting to ask what relevance these findings might have for syntactic theory.

First, these results provide us with some evidence that speaks to the issues of transparency. Rapid or immediate alignment between the grammar and the parser is at the very least consistent with the transparency thesis. In this spirit, subsequent work using the mismatch paradigm in eye-tracking has leveraged this close alignment between online parsing processes and grammatical constraints to investigate long-held distinctions made by formal linguistic theories. One important distinction is between co-argument and non co-argument anaphors/reflexives. As we will see, this long-held grammatical distinction between argument and co-argument anaphors does not map cleanly onto online measures. If we adopt a tight link between the grammar and the parser, it stands to reason that these results can inform inquiry into the (grammatical) status of the coargumenthood distinction.

Since the influential work of Pollard & Sag (1992), it has been suggested that only reflexives whose binder is a coargument of the same syntactic or semantic predicate are strictly subject to Principle A of the Binding Theory; those with non-coargument antecedents are exempt anaphors that can participate in anaphoric relations that violate Principle A (Pollard & Sag, 1992; see also the notion of a 'logophor' in Reinhart & Reuland, 1993). Cunnings and Sturt (2014) use eye-tracking-while-reading to compare the processing of traditional and exempt reflexives using a gender-mismatch design. They investigated whether resolution of exempt reflexives was susceptible to interference from non-local NPs, with the goal of ascertaining if the structural position of the reflexive had any effect on the degree to which non-local antecedents might be considered. The first experiment looked at reflexive pronouns in direct object position (10). The second and third looked at reflexive pronouns inside picture NPs, without (11) or with (12) a local possessor, respectively. Picture noun phrase reflexives without a possessor are typically taken as parade-cases of exempt anaphors.

(10) Jonathan/Jennifer was walking through the military barracks.

Coargument reflexives

a. She/He heard that the soldier had positioned **himself** in the middle of the mess hall.
b. She/He heard that the soldier had positioned **herself** in the middle of the mess hall.

(11) Jonathan/Jennifer was walking through the military barracks.

Picture NP reflexives without possessors

a. She/He heard that the soldier had a picture of **himself** in the middle of the mess hall.

b. She/He heard that the soldier had a picture of **herself** in the middle of the mess hall.

(12) Jonathan/Jennifer was walking through the military barracks.

Possessed Picture NPs

a. She/He heard about the soldier's picture of **himself** in the middle of the mess hall.

b. She/He heard about the soldier's picture of **herself** in the middle of the mess hall.

In end-of-sentence interpretation judgments, Cunnings and Sturt confirmed that PNP reflexives (and especially possessed picture noun phrases) were more likely to take a non-local antecedent, and more so when the local antecedent was a (stereotypical) gender mismatch. This finding was not mirrored in the eye movement record. In all of their experiments, Cunnings and Sturt observed the expected gender-mismatch effect either at the reflexive pronoun or in the spillover region: reading times were slower when the local antecedent mismatched the reflexive's features. In no experiment did they find evidence that the gender of the non-local antecedent influenced early processing of the reflexive. A cumulative progression analysis did reveal, however, that comprehenders were slower to process picture noun phrase reflexives than coargument reflexives or possessed picture noun phrase reflexives (see also Burkhardt, 2005).

These results suggest that the parser treats co-argument and picture noun phrase reflexives similarly in at least one respect: in early reading, comprehenders only seem to entertain local, Principle A antecedents for all types of reflexives in the configurations that Cunnings and Sturt tested (though see Kaiser, Runner, Sussman & Tanenhaus, 2009 Runner, Sussman & Tanenhaus, 2003; 2006 for contrasting evidence from the visual world paradigm).

Recent research suggests, however, that comprehenders may even be willing to entertain non-local antecedents for coargument reflexives under certain conditions. Parker and Phillips (2017) investigated sentences as in (13; from their Experiment 3):

(13) The talented actor/actress mentioned that …

a. the attractive spokesman praised **himself** for a great job.

b. the attractive spokeswoman praised **himself** for a great job.

c. the attractive spokeswomen praised **himself** for a great job.

Parker and Phillips' studies extended the mismatch paradigm from Sturt (2003) by including both 1-feature accessible mismatch conditions (13b) and 2-feature mismatch conditions (13c). Across three experiments, they consistently observed that reflexives in sentences like (13c) were read more quickly when the non-local antecedent matched the reflexive's features; important, this effect was evident even in some early measures. Parker and

Phillips suggested that sensitivity to the non-local antecedent was the consequence of an antecedent retrieval mechanism that jointly considers structural and morphosyntactic information in selecting an antecedent (e.g. Lewis & Vasishth, 2005; Lewis, Vasishth & van Dyke, 2006; McElree, 2006; *i.a.*). On this view, though structural cues often carry the day (creating the appearance of strong Principle A sensitivity in many contexts), when the local antecedent is a very poor match to the reflexive's features, the non-local antecedent may sometimes be retrieved and considered.

Parker and Phillip's results show that sensitivity to non-local antecedents can arise in comprehension due to properties of the memory access mechanisms that mediate the online formation of reflexive-antecedent dependencies. However, Sloggett (2017) showed that the likelihood of accessing a non-local antecedent for reflexives is also conditioned on whether that antecedent is a tempting perspective center for an utterance (see Culy, 1994; Kuno, 1972; Sells, 1987; Sloggett, 2017). In particular, Sloggett observed that when the non-local antecedent in examples like (13) is a source of information, comprehenders access it more readily in early processing than when it is a receiver of information (Sloggett, 2017; see also Kaiser et al., 2009, for sensitivity to sourcehood for picture noun phrase reflexives). Similarly, Sloggett (2017) observed that access to non-local antecedents was impeded when the local subject is a tempting perspective center (e.g. an indexical pronoun such as *I*). Sloggett (2017) hypothesizes that sensitivity to non-local antecedents is mediated through an encoding of the logophoric center of an utterance, and that comprehenders can access this encoding more readily the reflexive's retrieval cues are a poor match to the local subject.

The results of Parker and Phillips, and Sloggett, provide an interesting complement to Cunnings and Sturt. While some theoretical accounts (Pollard & Sag, 1992; Reinhart & Reuland, 1993; van Valin & La Polla, 1997) have focused on the co-argument / non co-argument distinction as a critical grammatical factor in controlling access to non-local antecedents, the on-line eye-tracking data suggest a different picture. Co-argument and non co-argument reflexives show a similar sensitivity to local antecedents when directly compared (Cunnings and Sturt, 2014). However, readers do access non-local antecedents for co-argument reflexives in online processing in certain contexts, as a function of feature match to the Principle A antecedents (Parker & Phillips, 2017), and as a function of their sensitivity to perspectival elements (Sloggett, 2017). The eye-tracking data seem to imply that co-argument and non co-argument reflexives are more similar than different, and instead, processing factors (feature match to the local subject) and other grammatical factors (discourse roles of potential antecedents) have a greater role to play in controlling access to non-local antecedents.

If one maintains a strong relationship between the parser and the grammar, then these results seem to call into question theoretical distinctions drawn on the basis of intuitive data. In this respect, they align with other results from visual world methodologies (Kaiser et al., 2009; see especially discussion in Runner et al., 2003, 2006; Runner & Head, 2014). It is of course possible that the divisions among anaphors suggested by the online data do not, in fact, align with grammatically active divisions. However, in an interesting point of convergence, a similar conclusion is reached by Isabelle Charnavel and colleagues on the basis of acceptability judgment data (Charnavel & Sportiche, 2016).

6 Conclusions

This chapter was intended to provide aspiring experimental syntacticians with an introduction to and rough guide on how to answer questions about the grammar using eye-tracking-while-reading. We offered basic information on the experimental paradigm, an overview of behavioral models of reading, identified points of divergence between various models of the parser, and discussed challenges in deciding on appropriate mapping hypotheses between grammatical knowledge and parsing performance. We believe that experimental work that lacks a clear understanding of these issues will have only a limited impact on the broader experimental syntax enterprise, but work that succeeds in balancing the methodological and theoretical issues has a good chance of making important contributions to our understanding of human syntactic competence.

References

Alcocer, P. & Phillips, C. (2012). Using relational constraints in content-addressable memory architectures for sentence parsing. Unpublished manuscript.

Arnett, N., & Wagers, M. (2017). Subject encodings and retrieval interference. *Journal of Memory and Language*, *93*, 22-54.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of experimental psychology: Learning, memory, and cognition*, *28*, 748-769.

Barker, C. (2012). Quantificational binding does not require c-command. *Linguistic inquiry*, 43, 614-633.

Berwick, R., & Weinberg, A. (1984). The grammatical basis of linguistic performance: Language use and language acquisition. Cambridge, MA: MIT Press.

Bicknell, K., Higgins, E., Levy, R., & Rayner, K. (2013). Evidence for cognitively controlled saccade targeting in reading. In M. Knauff, M. Pauen, N. Sebanz and I. Wachsmuth (eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 197-202).

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras and C. Clifton Jr. (eds), *The on-line study of sentence comprehension: Eye-tracking, ERP, and beyond* (pp. 51-76). East Sussex, UK: Psychology Press.

Boland, J. E. (2005). Cognitive mechanisms and syntactic theory: Arguments against adjuncts in the lexicon. In Cutler, A.E. (ed.), Twenty-first Century Psycholinguistics: Four Cornerstones (pp 23-42). Erlbaum UK.

Boston, Marisa Ferrara, John T. Hale, Reinhold Kliegl, Umesh Patil, and Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research,* 2, 1-12.

Burkhardt, P. (2005). *The syntax-discourse interface: Representing and interpreting dependency* (*Vol. 80*). Philadelphia, PA: John Benjamins Publishing.

Carminati, M. N., Frazier, L., & Rayner, K. (2002). Bound variables and c-command. *Journal of semantics*, 19, 1-34.

Charnavel, I., & Sportiche, D. (2016). Anaphor binding: What French inanimate anaphors show. *Linguistic inquiry*, 47, 35-38.

Chomsky, N. (1965). Aspects of the theory of syntax. *Cambridge, MA: MIT Press*.

Chomsky, N. (2005). Three factors in language design. *Linguistic inquiry*, 36, 1-22.

Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language*, *86*, 1-19.

Clifton, C., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, 2, 234-250.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. Fischer, W. Murray and R. Hill (eds.), *Eye movements: A window on mind and brain* (pp. 341-372). Amsterdam, the Netherlands: Elsevier.

Clifton Jr, C. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of experimental psychology: Learning, memory, and cognition*, *39*, 487-501.

Culy, C. (1994). Aspects of logophoric marking. *Linguistics, 32,* 1055-1094.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of memory and language*, *75*, 117-139.

Cunnings, I., & Sturt, P. (2017). Locality and discourse constraints in reflexive processing. Poster presented at the 30th annual CUNY Human Sentence Processing Conference (Boston, MA).

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive linguistics*, *2*(1), 1-63.

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*, 200-224.

Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. *Language and Linguistics Compass*, 8, 171-187.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85-103.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. Journal of Memory and Language, 27, 429–446.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 20, 641–655.

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

Engdahl, E. (1983). Parasitic gaps. *Linguistics & Philosophy.* 6, 5–34.

Engdahl, E. (1997). Relative clause extractions in context. *Working papers in Scandinavian syntax*, *60*, 51-79.

Engelmann, F., Vasishth, S., Engbert, R., & Kliegl, R. (2013). A framework for modeling the interaction of syntactic processing and eye movement control. *Topics in cognitive science*, *5*, 452-474.

Ferreira, F. & Henderson, J. (2004). *The Interface of Language, Vision, and Action.* East Sussex, UK: Psychology Press.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1, 71-83.

Frazier, L. (2008). Processing ellipsis: A processing solution to the undergeneration problem. In *Proceedings of the 26th West Coast Conference on Formal Linguistics* (pp. 21-32). Somerville, MA: Cascadilla 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.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and cognitive processes*, *8*(4), 573-633.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*.

Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in cognitive sciences* 14, 233–234.

Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and cognitive processes*, *28*, 88-124.

Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*, 82, 147-161.

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.

van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. *Reading as a perceptual process*, 62, 1-648.

van Gompel, R. P., 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 Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 52, 284-307.

Gorrell, P. (1993). Evaluating the direct association hypothesis: A reply to Pickering and Barry (1991). *Language and Cognitive Processes*, 8, 129-146.

Grant, M., Dillon, B. & Sloggett, S. (submitted). Ambiguity resolution in attachment and pronominal reference.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Association for Computational Linguistics.

Hare, M., K. McRae, and J. L. Elman. 2003. Sense and structure: meaning as a determinant of verb subcategorization preferences. *Journal of memory and language*, 48, 281–303.

Huey, E. B. (1908/1968). The psychology and pedagogy of reading. New York: McMillan (Reprinted: Cambridge, MA: MIT Press, 1968).

Jaeger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of memory and language*, *94*, 316-339.

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. Journal of Experimental Psychology: Human Perception and Performance, 34, 1560–1579.

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, *111*, 228-238.

Just M. A. & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. Kieras and M. Just (Eds.), *New methods in reading comprehension research* (pp. 151–182). Hillsdale, NJ: Erlbaum.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55-80.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*, 239-261.

Kreiner, H., Garrod, S., & Sturt, P. (2013). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes*, *28*, 829-874.

Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic inquiry*, 3, 269-320.

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing.* Unpublished PhD dissertation, University of Maryland, College Park.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: evidence from bound variable pronouns. *Journal of Memory and Language*, 82, 18-40.

Lasnik, H., Depiante, M. A., & Stepanov, A. (2000). Syntactic structures revisited: Contemporary lectures on classic transformational theory (Vol. 33). Cambridge, MA: MIT Press.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of psycholinguistic research*, *44*, 27-46.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, 10, 447-454.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements while reading disappearing text: Is there a gap effect in reading? *Vision research*, 44, 1013–1024.

Maling, J., & Zaenen, A. (1982). A phrase structure account of Scandinavian extraction phenomena. In *The nature of syntactic representation* (pp. 229-282). Springer, Dordrecht.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language*, 94, 119-133.

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis?. *Journal of memory and language*, *65*(2), 109-127.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and cognitive processes*, 28, 1545-1578.

Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *The linguistic review*, 22(2-4), 429-445.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.

McConkie, G. W., & Yang, S. N. (2003). How cognition affects eye movements during reading. In J. Hyona, R. Radach, & H. Deubel (Eds.), The mind's eye: Cognitive and applied aspects of eye movement research (pp. 413–427). Amsterdam: North-Holland Press.

McElree, B. (2006). Accessing recent events. In B. Ross (Ed.), *Psychology of learning and motivation, vol. 46* (pp. 155–200). San Diego, CA: Elsevier Academic Press.

Moulton, K. & Han, C-H. (to appear). Scope vs. C-Command: An Experimental Investigation. *Language.*

Nicenboim, B., Engelmann, F., Suckow, K., & Vasishth, S. (m.s.). Number interference in German: Evidence for cue-based retrieval. Psyarxiv e-print. Retrieved from: http://www.ling.uni-potsdam.de/~nicenboim/papers/NicenboimEtAl2016Number.pdf

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of psycholinguistic research*, *18*, 5-19.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in psychology*, *7*.

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of memory and language*, 94, 272-290.

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

Phillips, C. (1996). *Order and structure*. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82: 795-823.

Phillips, C. (2010). Should we impeach armchair linguists. *Japanese/Korean linguistics*, 15, 49-64.

Phillips, Colin. (2013). 'On the nature of island constraints I: language processing and reductionist accounts.' In Jon Sprouse and Norbert Hornstein (eds), *Experimental syntax and island effects*, Cambridge: Cambridge University Press, 64-108.

Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. *Oxford handbook of psycholinguistics*, pp. 739-756. New York, NY: Oxford University Press.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and cognitive processes*, 6, 229-259.

Pollatsek, A. & Treiman, R. (2015). *The Oxford handbook of reading*. New York, NY: Oxford University Press.

Pollard, C., & Sag, I. A. (1992). Anaphors in English and the scope of binding theory. *Linguistic inquiry*, *23*, 261-303.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. *New methods in reading comprehension research*, *118*, 91-118.

Rayner, K. (1978). Eye movements in reading and information processing. Psychological Bulletin, 85, 618–660.

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

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. Journal of Experimental Psychology: Human Perception and Performance, 30, 720–732.

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., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. Psychological Science, 14, 385–389.

Rayner, K., & Pollatsek, A. (1989). The psychology of reading. New York: Erlbaum (reprinted 1995, Erlbaum).

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*, SI21-SI49.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 1290–1301.

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. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. Psychonomic Bulletin & Review, 16, 1–21.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic inquiry*, *24*, 657-720.

Runner, J. T., & Head, K. D. (2014). What can visual world eye-tracking tell us about the binding theory?. *Empirical issues in syntax and semantics*, *10*, 269-286.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition*, *89*, B1-B13.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrase. *Cognitive science*, *30*, 193-241.

Safir, K. (2004). The syntax of anaphora. New York, NY: Oxford University Press.

Scheepers, C. Hemforth, B., Konieczny, L. & van Gompel, R. P. G. (unpublished manuscript). Monotonicity in head-final sentence processing: Top–down prediction of verb valency.

Schotter, E. R., & Rayner, K. (2015). The work of the eyes during reading. In A. Pollatsek and R. Treiman (eds.), *The Oxford handbook of reading* (pp. 44-59). New York, NY: Oxford University Press.

Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry*, *18*(3), 445-479.

Sloggett, S. (2017). *When errors aren't: How comprehenders selectively violate Binding Theory*. Unpublished Ph.D. Dissertation, University of Massachusetts, Amherst.

Sprouse, J. & Almeida, D. (2013). The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and cognitive processes,* 28, 222-228.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311-327.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and cognitive processes*, *1*(3), 227-245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of memory and language*, *48*, 542-562.

Townsend, D. J., & Bever, T. G. (2001). Sentence comprehension: The integration of habits and rules. Cambridge, MA: MIT Press.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454-475.

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.

Trotzke, A., Bader, M., & Frazier, L. (2013). Third factors and the performance interface in language design. *Biolinguistics*, *7*, 1-34.

Tutunjian, D., Heinat, F., Klingvall, E., & Wiklund, A. L. (2017). Processing relative clause extractions in Swedish. *Frontiers in psychology*, *8*, 2118.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*, 685-712.

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley interdisciplinary reviews: Cognitive science*, *4*, 125-134.

Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge University Press.

Vosse, T., & Kempen, G. (2009). In defense of competition during syntactic ambiguity resolution. *Journal of psycholinguistic research*, *38*, 1.

Wagers, M. W. (2013). Memory mechanisms for wh-dependency formation and their implications for islandhood. ' In Jon Sprouse and Norbert Hornstein (eds), *Experimental syntax and island effects*, Cambridge: Cambridge University Press, 161-85.