

Active prediction of syntactic information during sentence processing*

Zeynep Ilkin

ZEYNEPILK@GMAIL.COM

Psychology

University of Edinburgh

7 George Square

Edinburgh EH8 9JZ

Patrick Sturt

PATRICK.STURT@ED.AC.UK

Psychology

University of Edinburgh

7 George Square

Edinburgh EH8 9JZ

Editor:

Abstract

We describe an eye-tracking experiment that tested the effect of syntactic predictability on skipping rates during reading. We found that plural noun phrases were skipped more often than singular noun phrases, in syntactic contexts which induced a high expectation for a plural. We interpret this effect as evidence that the plural noun phrase has been predicted ahead of time. The results indicate that the examination of skipping rates might be a useful tool for the investigation of syntactic prediction effects.

Keywords: Incrementality, Parsing, Eye-tracking, Reading, Prediction

1. Introduction

Successful language comprehension requires the incremental integration of partial linguistic information. For example, in dialogue, part of an utterance may be inaudible due to noise in the environment; a speaker's utterance may stop (or be continued by an interlocutor) before it is formally complete; or a speaker may re-start an utterance mid-stream. In order for communication to be successful in these situations, the processing mechanism must be extremely flexible: incomplete pieces of structure may have to be integrated in order to create partial interpretations, or missing input items may have to be inferred from the information available.

In contrast, previous theories of human parsing have often lacked this type of flexibility, due to the adoption of bottom-up processing strategies. Such theories assume that the basic building blocks of interpretation are complete syntactic constituents, or that the building of structure requires the presence of a licensing phrasal head in the input (Mulders, 2002; Pritchett, 1991). The prediction is that there is a limit to the degree to which partial linguistic information can be used, because some crucial part of a constituent, for example, its lexical head, needs to be recognised before that

*. This research was presented at the CUNY sentence processing conference, in Davis, CA, 2009; at the ZiF workshop on Incrementality and Verbal Interaction, in Bielefeld, Germany, 2009; and at the European Conference on Eye Movements, in Southampton, UK, 2009. We thank the audiences of these conferences, as well as three anonymous reviewers, for their comments.

constituent can be integrated into the previous syntactic context. Such theories would require extra mechanisms to explain the ease with which partial utterances are integrated in dialogue, and they are also incompatible with empirical evidence in reading research. For example, evidence from the processing of Japanese, a verb final language, shows that long-distance dependencies can be formed into a subordinate clause before the verbal head of that clause is recognised in the input (Aoshima et al., 2004), and analogous effects have been found in English (Lee, 2004). These findings would not be expected if the bottom-up input of the verb were necessary for the formation of the dependency.

In contrast to the data-driven character of processing in bottom-up models, recent work has argued for a greater role for top-down prediction in human parsing. The idea is that comprehenders activate representations of possible continuations of the sentence, and these are constantly updated as each new word comes in (see, for example, Levy (2008) for a probabilistic model incorporating this idea). Predictive processing has an obvious relevance for explaining the flexibility of dialogue processing (Pickering and Garrod, 2007): some such process must underlie the ability of interlocutors to complete each other's utterances. Moreover, in cases where bottom-up information is unavailable due to noise in the environment, comprehenders may be forced to rely on top-down information.

Despite the intuitive appeal of predictive processing for modelling human linguistic performance, unambiguous experimental evidence for prediction has been surprisingly elusive. As theoretical processing models become more sophisticated, they begin to make fine-grained predictions about the time-course of prediction, and about the level of representation that is predicted. It then becomes increasingly important to develop experimental techniques that test prediction at such fine-grained levels. The problem is that, although a large body of work has demonstrated processing facilitation for words that are expected given the context, it is not always possible to interpret such results as evidence for the prediction of the word ahead of time. For example, studies have shown that lexical decision times are faster when the target word is predictable given the context than when it is not (relative to controls). Schwanenflugel and Shoben (1985) recorded lexical decision times for the final (underlined) word in visually presented stimuli such as the following, where the word was preceded by a separately presented context:

- (1)
 - a. **High constraint: expected word**
The worker was criticised by his boss
 - b. **Neutral baseline**
xxx xxxxxx xxx xxxxxxxxxxxx xx xxx boss
 - c. **High constraint: unexpected word**
The worker was criticised by his manager
 - d. **Neutral baseline**
xxx xxxxxx xxx xxxxxxxxxxxx xx xxx manager

Lexical decision responses to the expected word (*boss* in this case), were faster in the high constraint context relative to a baseline context consisting of a row of X's, while responses to the unexpected word (*manager*) did not differ from the baseline context. It is possible to interpret such findings in terms of predictive mechanisms. According to such an interpretation, the highly constraining context leads to the activation of the expected word's lexical entry ahead of time, thus allowing the recognition of the word to occur faster when it eventually appears in the input.

However, another possible interpretation of results such as these is that people build up a situation model based on the context sentence, and the expected word is simply easier to integrate into this situation model than the unexpected word (Traxler and Foss, 2000). Crucially, this explanation does not rely on the pre-activation of the test word, and thus the experiment does not afford unambiguous evidence of prediction.

Similar alternative interpretations can be made for demonstrations of predictive processing in other domains, such as syntax and reference. Van Gompel and Liversedge (2003) recorded participants' eye-movements while they read sentences containing cataphoric pronouns, such as the following:

- (2) a. **Congruent: masculine**
When he was fed up, the boy visited the girl very often.
- b. **Congruent feminine**
When she was fed up, the girl visited the boy very often.
- c. **Incongruent: masculine**
When she was fed up, the boy visited the girl very often.
- d. **Incongruent feminine**
When he was fed up, the girl visited the boy very often.

The stimuli always included a gender-marked pronoun in a preposed adverbial clause (*he* or *she*), which was either congruent or incongruent with the gender of the main clause subject (*the boy* or *the girl* in this case). Eye-movement measures showed that participants slowed down at or immediately after the main clause subject in the incongruent conditions relative to the congruent conditions. Van Gompel and Liversedge (2003) argued that the processor forms the referential dependency before the gender information of the main clause subject has been computed, resulting in a slow-down in the incongruent conditions, where the dependency has to be abandoned. However, such findings are also compatible with a highly predictive processing strategy, according to which the main clause subject position is predicted ahead of time, while the initial subordinate clause is being processed. The referential dependency is then formed between the cataphoric pronoun and this predicted subject position. This then causes processing difficulty in the incongruent conditions, due to the gender mismatch between the features of the predicted subject position and those of the head noun *boy/girl*. This type of predictive strategy has been suggested by Kazanina et al. (2007), and also by Kreiner et al. (2008), both of whom report similar effects. Kazanina et al. (2007) additionally showed that the difference between congruent and incongruent conditions disappears when the subject is ruled out as an antecedent position by binding principles, as in (3).

- (3) a. **Principle C**
It seemed worrisome to him that John/Ruth was gaining so much weight, but Matt didn't have the nerve to comment on it.
- b. **No constraint**
It seemed worrisome to his family that John/Ruth was gaining so much weight, but Matt didn't have the nerve to comment on it.

In (3a), Binding Principle C (Chomsky, 1981) rules out the subordinate clause subject (*John/Ruth*) as an antecedent of *him*, while this is not the case in (3b). Kazanina et al. (2007) demonstrated a

slow-down in self-paced reading times for the incongruent *Ruth* relative to the congruent *John* in (3b), but no such congruency effect was found in (3a), where the dependency was not licit in terms of binding theory. Kazanina et al. (2007) interpreted these results in terms of a predictive processing strategy which they term *active search*, in which the relevant position for the antecedent phrase is predicted ahead of time, but where the referential link is only made where this predicted phrase is a licit antecedent in terms of binding theory.

Kreiner et al. (2008) also argued for a predictive processing strategy on the basis of experimental data on cataphoric reference. In their experiment 2, Kreiner et al. (2008) recorded eye-movements while participants read sentences like those in (4):

- (4) a. **Stereotypical Gender: congruent**
After reminding himself about the letter, the minister immediately went to the meeting at the office.
- b. **Stereotypical Gender: incongruent**
After reminding herself about the letter, the minister immediately went to the meeting at the office.
- c. **Definitional Gender: congruent**
After reminding himself about the letter, the king immediately went to the meeting at the office.
- d. **Definitional Gender: incongruent**
After reminding herself about the letter, the king immediately went to the meeting at the office.

The design of Kreiner et al. (2008) incorporated both stereotypical gender-biased nouns (e.g. *minister* in (4a,b) is likely to be a man, but does not have to be), and also definitional gender nouns (e.g. *king* in (4c,d) has to denote a male by definition). The stereotypical or definitional gender could either be congruent or incongruent with a preceding reflexive, which had to be co-referential with the main clause subject. Kreiner et al. (2008) found a congruency effect for the definitional gender conditions, with longer reading times immediately following *king* in (4d) than in (4c). However, no such difference was found between the two stereotype gender conditions (4a) and (4b). Kreiner et al. (2008) interpreted this result in terms of a predictive processing strategy in which the dependency between *himself/herself* and the main clause subject was formed in advance, constraining the gender of the subject before its appearance in the input. The congruency effect for the definitional nouns can be explained on the assumption that the gender information for these nouns is retrieved at lexical access, and this causes a slow-down due to gender mismatch in the incongruent condition. The lack of such a congruency effect for the stereotypical nouns can be explained on the assumption that the gender information for these nouns is determined through fit with the context, or via stereotype inference, rather than through the retrieval of a gender feature at lexical access. The predictive dependency formation strategy results in the gender of the main clause subject being determined before the stereotype role name is processed in the input. Since the gender is already available, there is no need for a stereotype inference, and the lack of mismatch effect for (4a) and (4b) can be explained.

However, although the notion of predictive dependency formation is compatible with the results of both Kazanina et al. (2007) and Kreiner et al. (2008), this is not the only possible explanation. In both cases, it is possible that the referential dependency was formed only after the head noun

of the antecedent noun phrase had been processed in the input. The congruency effect reported by Kazanina et al. (2007) (see (3) above) could have arisen if the parser did not predict the position of the antecedent of the cataphoric pronoun in advance. According to this non-predictive account, the processor would form the referential dependency when the critical word of the antecedent phrase (*John/Ruth*) has been processed. For example, the processor might attempt to update co-reference relations whenever any potentially referential noun phrase is read. It might be at this point that the structural details relevant to binding constraints are taken into account, and referential dependencies are considered only where allowed by binding theory. On this account, the gender congruency effect that Kazanina et al. (2007) obtained could be explained in a number of ways; for example, as suggested by Van Gompel and Liversedge (2003), it might be the case that the referential dependency is formed after the part-of-speech information of the antecedent is retrieved, but before the gender information has been retrieved, due to architectural constraints on the timecourse of the availability of linguistic information. This would lead to the processing difficulty in the incongruent condition, since the dependency is formed before the gender is checked. An alternative possibility is that the processing difficulty arises from competition among simultaneously active constraints, as in constraint satisfaction models of parsing (McRae et al., 1998); in the case of the incongruent condition, the gender constraint would lead to a bias against the referential dependency, while other constraints, such as, for example, the subject preference for pronoun antecedents would lead to a bias in favour of the referential dependency. The competition between these constraints would explain the processing difficulty.

The results of Kreiner et al. (2008) (see example (4) above) could also be explained in terms of a non-predictive processing strategy. Again, according to this account, it is assumed that the processing of the potential referential antecedent (in this case, the main clause subject) triggers an update of co-reference information. It is at this point that the processor registers the possibility of a co-reference relation between the cataphoric reflexive and the main clause subject (in fact, in this case, the co-reference is grammatically obligatory). The congruency effect for the definitional gender nouns can be explained as before, if we assume either an architectural constraint on the temporal order of linguistic information, or alternatively, competition among simultaneously active constraints. The lack of a congruency effect for the stereotype nouns could be explained if the referential dependency is formed, and the gender of the main clause subject determined, before any attempt to execute the stereotype gender inference.

The above literature review has examined three examples of effects that have been interpreted in terms of predictive processing, where alternative non-predictive accounts are possible. The common factor in all of these studies, as well as many others in the literature, is that the experimental manipulation allows processing differences to be observed only at, or following, the predicted element. For example Schwanenflugel and Shoben (1985) tested prediction by measuring lexical decision times to the predicted word. Similarly, Kazanina et al. (2007) and Kreiner et al. (2008) tested prediction by measuring reading times on a phrase whose gender feature had been putatively predicted. In all such cases, the evidence for prediction consists of facilitation of the processing of the predicted element and/or processing difficulty for the unpredicted element.

There are at least two different ways in which one can make a more watertight case for prediction. The first is an argument based on the speed with which the relevant effects are observed. For example, Lau et al. (2006) conducted an Event Related Potential (ERP) study in which an ELAN effect (Early Left Anterior Negativity) was elicited for a (normally) ungrammatical sequence consisting of a possessor followed by the word *of* (e.g. ... *Max's of* ...). Lau et al. (2006) manipulated

whether or not the sequence was licensed by the possibility of ellipsis (e.g. *Although Erica kissed Mary's mother, she did not kiss Dana's ...*). The results showed that the licit and illicit conditions began to diverge at around 200 msec after the onset of the critical stimulus. Taking into account previous estimates of the timecourse of lexical access at around 100-200 msec (Sereno and Rayner, 2003), the authors argued that the early occurrence of the ELAN effect was unlikely to have been observed unless the ellipsis had been predicted ahead of time.

A second type of evidence that has been used to argue for predictive processing has involved demonstrations that the predicted element affects processing *before* it has been processed in the input. In the following paragraphs, we will describe two recent lines of research which take this approach.

In an Event Related Potential (ERP) study, DeLong et al. (2005) examined the processing of sentences such as (5), with a word-by-word visual presentation:

- (5) a. The day was breezy so the boy went outside to fly a kite.
- b. The day was breezy so the boy went outside to fly an airplane.

In (5), the context induces a strong expectation for the word *kite*, as verified by a norming study reported by DeLong et al. (2005). This expectation is fulfilled in (5a), but not in (5b), where *airplane* is not the most strongly expected continuation. In the study, DeLong et al. (2005) sought to demonstrate that the specific word form *kite* had been predicted before its appearance in the input. The study exploited the fact that the form of the English indefinite differs according to whether the following word begins with a vowel (*an*) or a consonant (*a*). In (5), the expected word *kite* begins with a consonant, which is consistent with the definite article *a*, but inconsistent with the form *an*. The ERP analysis focused on responses to the indefinite article. At the indefinite article, a greater negative deflection was found for *an* in (5b) than for *a* in (5a), and this effect was consistent with the N400 ERP component, which is known to be sensitive to the degree of contextual expectation of a word. Moreover, the authors demonstrated that the size of this N400 effect correlated reliably with the degree of expectation as measured by the prior norming results. Thus, the authors found clear evidence of anticipation of a specific word *before* that word had been encountered in the input.

Using a similar experimental logic, Van Berkum et al. (2005) examined the ERP responses to Dutch inflected pre-nominal adjectives as a function of the degree of expectation of an up-coming noun. They found a positive deflection in conditions where the pre-nominal adjective did not agree in gender with the predicted noun, relative to conditions where the adjective agreed with the predicted noun. This is therefore also evidence that the prediction is active before the predicted element is encountered in the input.

A second line of research showing anticipation before the predicted element comes from the visual world paradigm. In this technique, participants look at a depicted scene while they listen to a spoken sentence. The relative proportions of looks to various target and distractor objects in the depicted scene are analysed over time, as a function of manipulations in the spoken sentence. Altmann and Kamide (1999) showed that people began to look at objects that would be expected given the sentential context, even before the relevant object was mentioned in the spoken sentence. For example, in a sentence like *The boy will eat the cake*, participants' looks to a depicted cake began to increase soon around the offset of the word *eat*, relative to a condition where the spoken sentence did not lead to the expectation of the word *cake* (i.e. *The boy will move the cake*). This experiment, as well as others using the visual world paradigm (Kamide et al., 2003) have shown that people can use linguistic knowledge to anticipate the mention of visually presented objects ahead of

time. However, it is currently not known whether these types of effects rely on the presence of the visually presented objects in the scene. The presentation of a small number of potential referents at the start of each trial might effectively narrow down the possibilities for anticipation to a degree that allows the anticipatory effects to be observed, but such effects might not occur when the visual objects are not present.

To summarise, the above literature review has highlighted some of the difficulties involved in interpreting experimental evidence for prediction in sentence processing. The most convincing evidence comes from studies in which effects of prediction are demonstrated before the predicted element is encountered in the string. It is also possible, given assumptions about the time-course of processes involved, to argue for predictive mechanisms based on the early appearance of congruency effects, as argued by Lau et al. (2006).

In the experiment reported below, we examine the effect of cataphoric pronouns on the prediction of features of their antecedents, as did Kreiner et al. (2008) and Kazanina et al. (2007). However, we use a new source of evidence, namely word-skipping in reading, to examine the prediction of syntactic information in sentence processing. We argue that the use of this measure makes a more convincing case for predictive effects in parsing.

2. Experiment

The experiment used eye-tracking in reading to test the extent to which people maintained expectations of plural morphological information on a noun phrase before the phrase had been read. As well as the usual eye-movement measures, which involve measurement of fixation time on particular words or phrases of interest, here we make crucial use of the measurement of *skipping* rates as a measure of predictability. In eye-movement research, the term “skipping” refers to the phenomenon whereby a word, or a sequence of words is not fixated directly during initial reading, but instead, the reader makes an eye-movement that “skips” over the word. The “skipping” eye-movement is launched from material preceding the critical word, and it lands on material that follows the critical word. Around one in three words are skipped in normal reading in English (Brysbaert et al., 2005).

Previous studies have already established a link between lexical predictability and word skipping (Balota et al., 1985; Rayner et al., 2004). For example, Rayner et al. (2004) recorded eye-movements while participants read sentences like (6):

- (6) a. **Predictable:**
Before warming the milk, the babysitter took the infant’s *bottle* out of the travel bag.
- b. **Unpredictable:**
To prevent a mess, the caregiver checked the infant’s *bottle* before leaving.

Off-line tests established that the word *bottle* was a very frequent continuation of the sentence prefix in (6a), and that it was a very rare continuation in (6b). In the eye-tracking experiment, readers skipped over the critical word *bottle* more often in the predictable condition than in the unpredictable condition. In other words, the proportion of trials in which *bottle* received a fixation during initial reading was greater in the unpredictable condition than the predictable condition. In the E-Z reader model of eye-movement control (Reichle et al., 2003) skipping occurs in cases where, during a fixation on word N, attention shifts to word N+1¹, and lexical processing for N+1

1. Word N+1 refers to the word immediately to the right of word N.

is initiated. If the first stage of lexical processing of $N+1$ is completed quickly enough, an eye-movement is programmed to word $N+2$, and the previously programmed eye-movement to word $N+1$ is cancelled. The eye-movement launched from word N to word $N+2$ results in the skipping of word $N+1$. In the study of Rayner et al. (2004), the higher skipping rates for the predictable condition relative to the unpredictable condition can be explained if we assume that the critical word is partially activated ahead of time in the predictable condition, due to the contextual constraint. When the reader is fixating the word immediately preceding the critical word *bottle*, the increased activation of *bottle* allows the first stage of lexical processing of this word to be completed relatively quickly, increasing the probability of skipping the word relative to the unpredicted condition.

In the current experiment, we used a similar way to measure predictability with eye-movements, except that in this case, we measure the prediction of a syntactic feature, namely a plural number feature on a phrase, rather than the prediction of a specific word.

The stimuli were designed to give a very high expectation for a plural noun phrase. This was done using cleft sentences with reciprocal anaphors, as in (7):

- (7)
- a. It was to each other that the girls from the school said that the children from next door wanted us to give advice.
 - b. It was to each other that the girl from the school said that the children from next door wanted us to give advice.

Due to binding constraints, the reciprocal *each other* requires a locally c-commanding plural noun phrase as its antecedent. In the cleft construction in (7), this requirement is satisfied via an unbounded dependency, with *each other* as the dependent element. We assumed that this would lead to a strong expectation for a plural subject of the *that*-clause, as in *the girls* in (7a). However, due to the fact that this requirement is licensed via an unbounded dependency, the plural antecedent may be more deeply embedded in the structure, and the subject of the *that*-clause is not grammatically constrained to be the antecedent. Thus, it is grammatically possible for the subject of the *that*-clause to be plural, as in (7a), or singular, as in (7b). In both cases, the globally correct antecedent is found more deeply embedded in the sentence (*us*). Although both (7a) and (7b) are both grammatical, only (7a) satisfies the strong expectation for the subject of the *that*-clause to be plural.

Thus, we expected (7b) to elicit processing difficulty relative to (7a). However, as we have seen in the review above, such evidence of processing difficulty at or following the predicted element would not necessarily imply that the plural feature of the subject is predicted ahead of time.

In this experiment, we therefore measured eye-movement behaviour that was initiated *before* the critical noun phrase is fixated in the string, in addition to the standard eye-movement measures based on reading time. Specifically, we measured the proportion of trials in which readers skipped the phrase *the girls* in initial reading, without fixating it.

If the skipping rates for *the girl(s)* vary by condition, this can be interpreted as evidence that some relevant information from this phrase has been processed by the cognitive system parafoveally, without the reader fixating it. Given that any skipping of the phrase *the girls* must have been launched from a fixation position earlier than the determiner *the*, and also given that the two conditions differ only in the morphology of the second word of the phrase (in this case, the presence vs. absence of the plural marker “s”), any such result must mean that this morphological information has been processed while the reader is fixating a position that is at least two words to the left of the predicted head noun. In cases where the number feature is highly predictable, we assume that the plurality of the subject noun phrase is computed ahead of time, facilitating the parafoveal processing

of the phrase, and increasing the probability that it is skipped. We will return to the mechanisms by which skipping might be affected by syntactic parsing mechanisms in the general discussion.

A second aim of the experiment was to assess the role of pre-verbal dependency formation in English. Recall that Aoshima et al. (2004) showed evidence for long distance dependency formation in advance of the verb in Japanese sentence processing, which they interpreted as evidence against the head-driven strategy for Japanese. The current experiment allows us to conduct a similar test for English, which, unlike Japanese, has a predominantly head-initial constituent order. Specifically, the experiment allows us to test whether the dependency between *each other* and *the girls* in (7) is made in advance of the verb *said*. This result would imply considerable postulation of structure in advance of bottom-up evidence, given that the dependency between *each other* and *the girls* requires (a) an appropriate binding configuration, and (b) an unbounded dependency, in which *to each other* is associated with an underlying indirect object position.

2.1 Method

2.1.1 PARTICIPANTS

Thirty two native speakers of English who were members of the University of Edinburgh community participated to the experiment. They were each paid £4 to participate in the experiment. All had normal or corrected to normal vision, and all were naïve to the purpose of the experiment.

2.1.2 MATERIALS

Experimental materials consist of 28 sets of sentences like (8) in a 2×2 factorial design, orthogonally manipulating the form of the initial PP (conjunction vs. reciprocal) and the number of the subject of the *that*-clause. For simplicity of exposition, we will refer to the reciprocal plural and singular conditions as *match* and *mismatch* respectively. We will refer to the two experimental factors as *PP-type* and *Number* respectively.

- (8)
- a. **Reciprocal: Plural (Match)**
It was to each other that the girls from the school said that the children from next door wanted us to give advice. The children liked us.
 - b. **Reciprocal: Singular (Mismatch)**
It was to each other that the girl from the school said that the children from next door wanted us to give advice. The children liked us.
 - c. **Conjunct: Plural**
It was to John and Mary that the girls from the school said that the children from next door wanted us to give advice. The children liked us.
 - d. **Conjunct: Singular**
It was to John and Mary that the girl from the school said that the children from next door wanted us to give advice. The children liked us.

The two conjunct conditions were included in the design for experimental control. They incorporated a conjoined noun phrase instead of a reciprocal. The reason for this is that the skipping rates for a plural noun like *girls* might differ from those for a singular noun like *girl*, irrespective of whether or not the noun has been predicted, for example, due to length differences. The conjoined conditions allow us to measure the skipping rates for exactly the same critical nouns (*girl/girls*), but

without the highly predictive context. Moreover, it is also possible that the fact that the conjoined phrase is semantically plural could control for the possibility that skipping rates might be affected by low-level priming of number information; for example, it might be the case that *each other* primes a plural noun simply because both elements bear a plural feature, and not because the plural noun is predicted.² Therefore, the conjoined conditions give the required degree of experimental control in order to observe a prediction-related effect on skipping rates. This effect should show up as an interaction, such that plural noun phrases are skipped more often than singular noun phrases when preceded by *each other* than when preceded by the conjoined phrase, but this effect should be absent, or reversed for the singular nouns.

All experimental items were followed by a second short sentence to prevent wrap up effects at the end of the first sentence. The 28 sets of items were distributed among 4 lists in a Latin Square design. Each participant was randomly assigned to a list. 28 experimental sentences were intermixed with 84 filler items. Sixteen of the experimental items and 44 of the filler items were followed by a comprehension question. All questions were answerable as ‘yes’ or ‘no’. The correct answers were counterbalanced, with half of the correct answers ‘yes’ and half ‘no’.

2.2 Results and Discussion

Prior to analysis, trials with track loss were eliminated. Fixations less than 80 msec in duration and within one character of the previous region were incorporated into the neighbouring fixation. Remaining fixations that were less than 80 msec were deleted.

In the following presentation of the results, we will give the results relating to word skipping first, and this will be followed by the results for standard reading-time based measures.

For the purposes of the data analysis, each experimental sentence was initially divided into the following regions:

Start of sentence	(<i>It was to</i>)
Anaphor	(<i>each other</i>)
Complementizer	(<i>that</i>)
First subject	(<i>the girls</i>)
First subject modifier	(<i>from the school</i>)
Main verb	(<i>said that</i>)
Second subject	(<i>the boys</i>)
Second subject modifier	(<i>from York</i>)
Embedded VP	(<i>requested us</i>)
End of sentence	(<i>to play piano</i>)

For the data analyses reported below, *Probability of first-pass fixation* is defined as the probability that the region received a fixation before subsequent regions were fixated. Note that this measure is directly related to skipping rates, since if a region is skipped, by definition, it cannot have received a first-pass fixation; and, conversely, if a region receives one or more first-pass fixation, it cannot have been skipped. Therefore, $\text{Prob}(\text{fixation}) = 1 - \text{Prob}(\text{skipping})$.

Table 1 gives the means and standard errors for the probability of initial fixation, for the first subject region.

2. However, this assumes that such priming is based on *semantic* number, as the nouns in the conjoined condition are each morphologically singular.

	First Subject (<i>the girl/girls</i>)
Probability of fixation (%)	
Match	87 (3.00)
Mismatch	96 (1.21)
Conj-Plur	94 (1.85)
Conj-Sing	89 (1.95)

Table 1: Probability of first-pass fixation for the first subject region (*the girl(s)*)

	Estimate	Std. Error	z value	p value
(Intercept)	3.01095	0.26115	11.529	< 2e-16 ***
Number	0.31602	0.28721	1.100	0.271198
PP-type	0.02925	0.28317	0.103	0.917743
Number×PP-type	1.99467	0.56662	3.520	0.000431 ***

Table 2: Results of statistical analysis for probability of fixation in the first subject region. key: “***”: $p < .001$. Model estimates are given in terms of the log-odds of the probability: $\log_odds(p) = \ln(\frac{p}{1-p})$

The measure of probability of fixation involves binary response data (either a trial received a first-pass fixation or it did not). Therefore, we use a logistic mixed effect regression for this measure (Jaeger 2008; see also Baayen et al. in press and Bates and Sarkar, 2007). The regression model included random intercepts for both subjects and items. The pattern of significance was not affected by the inclusion of extra parameters for random slopes, nor was model fit statistically improved by the inclusion of these parameters. We therefore report analyses for the simpler models, which include random intercepts only. Prior to analysis, the predictor variables were centered, such that categorical factors were transformed into numerical codes with a mean of zero and a range of 1. This procedure reduces collinearity between variables, and, in combination with sum coding of contrasts, also allows a straightforward interpretation of main effects and interactions in a way that is analogous to that of analysis of variance. The calculation of the significance of each effect is based on the Wald Z-test.

The results for the probability of one or more fixations in the first subject region are given in Table 2: The main result was a reliable interaction between PP-type and Number, such that there were fewer fixations (and therefore more skipping) on the plural when it was preceded by *each other* than when it was preceded by a conjoined phrase (87% vs. 94%), while the reverse effect was found for the singular phrase (96% vs. 89%). Pairwise comparisons revealed that both of these two contrasts were reliable (p 's < .05). Moreover, there were reliably fewer fixations on the plural NP than the singular NP when it was preceded by *each other* (p 's < .05), while the fixation rates for singular and plural NPs did not differ significantly when preceded by *and*, though the numerical pattern was in the opposite direction, with more fixations for the plural than the singular.

This pattern of data is exactly as expected if skipping rates are increased by the predictability of a plural NP. In the conjunct conditions, where there is no particular reason to predict a plural NP, fixation rates are statistically similar for singular and plural NPs, although the tendency is for more

fixations on the plural, which is expected due the extra length of the plural. The fact that there were differences from baseline for both the match and mismatch reciprocal conditions suggests that there may have been both facilitatory and inhibitory processes at work in this experiment. If this is the case, it would mean that the probability of skipping is not only increased (relative to baseline) when the number of the head noun matches the prediction, but also that it is *decreased*, again relative to baseline, when it mismatches the prediction; in other words, the mismatch leads to extra fixations on the critical noun phrase.

One potential objection to our interpretation of the skipping results is that the critical effect was based on only a very small number of skips over the course of the experiment. Clearly, as the very high rates of fixation show, a skip of a two word region is a very rare event, and most of our participants gave only one or two skips (or none at all) out of a maximum of seven in any given condition. Thus, the effect reported above is based on differences among a relatively small number of trials. It will therefore be important to replicate the present findings with the same design, but a much larger number of items. However, it is also true that, even though very few skips were being made in our experiment, the critical interactive pattern was quite robust over our participant sample. This is shown, firstly, by the fact that the fit of the logistic regression model was not improved by adding random slope parameters to allow the size of the critical interaction to vary by participant ($p > .9$). Moreover, we can gain an idea about the generality of the pattern by looking at the interaction for each individual participant. We defined an interaction score based on the proportions of fixations in each condition, which is then calculated individually for each participant: $[\text{prop}(\text{Match}) - \text{prop}(\text{Mismatch})] - [\text{prop}(\text{ConjPlur}) - \text{prop}(\text{ConjSing})]$. This score should have a negative value for participants who show the predicted interactive pattern, a positive value for participants who show an interactive pattern opposite to that predicted, and be zero for those who show no interactive pattern. It turns out that 20 of our 32 participants showed the predicted pattern, with only four showing the opposite pattern (binomial Sign test: $p = .0008$), and eight showing no interactive pattern (of whom six had ceiling levels of fixation, at 100% for all conditions). Thus, although the statistical pattern of the interaction is based on only a relatively small number of trials, the pattern was very general across participants, for those who did make two-word skips. However, the fact that six of our participants were at ceiling levels of fixation in all four conditions suggests that there may be individual differences in the propensity to skip two words at a time.

We now turn to the reading time measures. As demonstrated above, initial data analysis revealed that the first subject region (e.g. *the girls*) differed reliably by condition in the probability of first-pass fixation, making the interpretation of reading time measures hard to interpret. For both reading time measures (first pass and regression path), we therefore pooled the first and second subject regions with their corresponding modifiers (e.g. *the girls from the school*). These larger regions received equal rates of first pass fixation by condition, with all conditions receiving a first pass fixation 100% of the time in both regions. The reading time measures are defined as follows. First-pass reading time is the sum of all fixation durations from the first entry into the region from the left, until the first exit of the region, *either to the left or to the right*. Regression-path time is the sum of all fixation durations from the first entry into the region from the left, until the first exit of the region *to the right*. Note that regression path times may contain fixations outside the region following regressions out of the region. For both reading time measures, cases where the region is skipped in first-pass reading are treated as missing data, rather than contributing a zero value to the mean. As mentioned above, the fact that probabilities of initial fixations differed among conditions motivated the use of larger analysis regions for the reading time measures.

We analyzed the reading time measures starting from the anaphor region until the end of the initial sentence:

Anaphor	(<i>each other</i>)
Complementizer	(<i>that</i>)
First subject and the modifier	(<i>the girls from the school</i>)
Main verb	(<i>said that</i>)
Second subject and the modifier	(<i>the boys from York</i>)
Embedded VP	(<i>requested us</i>)
End of sentence	(<i>to play piano</i>)

Means and Standard Errors for both First-Pass Reading Times and Regression Path Times for each region are given in Table 3. Linear Mixed Effects Regression models were computed for each region and each measure. The t -statistic from these models is reported in Table 2.2. The Mixed Effects regressions included both experimental factors and their interaction as predictors, and also included random intercepts for subjects and items. Random slope parameters corresponding to main effects or the interaction were included only when justified by a significant improvement in model fit, based on a log-likelihood χ^2 test. Experimental variables were centered before analysis.

The main region we were interested in is the first subject plus modifier region. As mentioned above, reading time data for the first subject region alone (determiner plus noun) are hard to interpret due to skipping differences³. Here, if the reader is expecting to form the dependency as soon as possible, then in the Reciprocal Singular condition there should be longer fixation durations. The reader should be surprised to encounter a singular noun where a plural is expected. Also again in Reciprocal Singular Condition after a failure to form the dependency at the first possible subject position, if the parser continues to actively search for a possible subject NP to complete the dependency this should be visible in the following regions. The parser should initiate the dependency formation again at other possible subject positions, which could be interpreted as potential antecedents.

The results from the first 2 regions, the Anaphor region (*each other*) and the Complementizer region (*each that*) are not of theoretical interest in this experiment. The only significant difference between conditions was in Anaphor region, where the fixation durations was significantly longer when the sentences started with two conjoined proper names, compared with when it started with 'each other', both in First-Pass Reading Times (means, 367 vs. 312 ms.; respectively) and in Regression-Path Times (means, 493 vs. 352 ms.; respectively). This effect is probably due to differences in length and/or semantic complexity between the two conditions, and will not be discussed any further. There were no other significant differences in either measure for either the Anaphor or Complementizer regions. We will now discuss each following region separately for each measure.

The First Subject and the modifier (*the girls from the school*): For the First-Pass Reading Times there was no significant main effect of sentence initial PP or the number of the first subject NP. However the interaction between the PP and the number of the NP was significant. There were longer fixations when the sentence included a reciprocal and the NP was singular, compared to

3. For completeness, means are reported here for the first subject NP region (i.e. excluding modifier). Given the skipping differences, we give means excluding skipped trials followed by means including skipped trials as zero values (in parentheses) First pass: Match: 269ms (239ms); Mismatch: 263ms (252ms); Conj-Plural: 299ms (283ms); Conj-Singular: 277ms (252ms). Regression Path: Match 343ms (305ms); Mismatch: 321ms (306ms); And-Plural: 375ms (356ms); And-Singular: 421ms (380ms).

Measure	Anaphor	Comp.	NP1+Mod.	Main Verb	NP2+Mod.	Emb. VP	End of Sent.
FPT							
Match	312 (15)	225 (12)	599 (34)	361 (21)	697 (37)	338 (24)	738 (36)
Mismatch	311 (17)	242 (17)	632 (30)	349 (16)	695 (40)	366 (23)	717 (39)
Conj-Plur	374 (22)	261 (16)	608 (28)	339 (17)	668 (36)	347 (21)	724 (37)
Conj-Sing	360 (17)	240 (15)	573 (37)	356 (19)	689(42)	334 (23)	716 (35)
RPT							
Match	385 (24)	261 (17)	741 (41)	424 (25)	1059 (62)	417 (44)	1054 (101)
Mismatch	402 (25)	303 (30)	775 (47)	451 (30)	1199 (80)	489 (42)	883 (71)
Conj-Plur	615 (37)	287 (17)	772 (41)	443 (45)	1074 (62)	438 (36)	875 (69)
Conj-Sing	601 (40)	293 (31)	805 (50)	339 (33)	1031 (54)	454 (44)	849 (60)

Table 3: Means and Standard Errors for the First-Pass Time and Regression-Path Time Measures for Each Region

Measure		Anaphor	Comp.	NP1 +Mod.	Main Verb	NP2 +Mod.	Emb. VP	End of Sent.
FPT								
PP type	$t =$	-3.07*	< 1	< 1	< 1	< 1	< 1	< 1
Number	$t =$	< 1	< 1	< 1	< 1	< 1	< 1	< 1
PP \times Number	$t =$	< 1	< 1	2.01*	< 1	< 1	1.36 ^{ns}	< 1
RPT								
PP type	$t =$	-5.88*	< 1	-1.24 ^{ns}	< 1	1.58 ^{ns}	< 1	1.54 ^{ns}
Number	$t =$	< 1	< 1	-1.32 ^{ns}	< 1	1.07 ^{ns}	1.27 ^{ns}	-1.29 ^{ns}
PP \times Number	$t =$	< 1	< 1	< 1	< 1	2.05*	< 1	-1.42 ^{ns}

Table 4: t -values based on Linear mixed effect model of reading time data. Key: “ns”: Non-significant; “*”: $p < .05$. The expression “< 1” is a shorthand for $|t| < 1$. Significance at $\alpha = .05$ is based on the criterion of $|t| > 2$.

when it included conjunction and the first NP was singular (means, 632 vs, 573 ms.; respectively: $p < .05$). However, there were no reliable differences for the plural NP's ($p > .1$).

In regression path times, there was no reliable effect of sentence initial PP type, number of the first subject NP or the interaction between them.

The results for first pass time are consistent with the skipping rates for the first subject, and with the interpretation that the readers were expecting to have a plural noun at the first subject region only when sentences started with a reciprocal. In such cases, the parser may have been actively constructing the dependency as soon as the reciprocal was encountered, with the expectation to complete the dependency at the first subject position.

For the main verb region (*said that*) there were no reliable effects of either of the two experimental factors, or of the interaction between them, either for the first pass time or for regression path time. It seems that the processing difficulty for the reciprocal singular condition (as observed in the first pass times) does not spill over to this region.

For the second subject NP and the modifier (*the boys from York*), in First-Pass Time, there were no reliable effects. However in Regression-Path Times, the interaction between the two factors was significant. The pairwise comparisons revealed that the Reciprocal-singular condition was read more slowly than the Conjunct Singular (mismatch) condition (Reciprocal-Singular, 1199 vs. Conjunct-Singular, 1031), $p < .05$, but there was no such corresponding effect for the plural conditions ($p > .1$). For the reciprocal singular condition, this region was the first region that would allow the parser to complete the unbounded dependency after an unsuccessful attempt at the First Subject NP region. The longer fixation durations for the regression path times could indicate that the parser is reattempting to form the dependency at this region. Considering that the regression path time measure includes fixations that are made to the left of the region, the longer fixation durations could indicate the reactivation of the sentence initial PP and its attempted integration with the second subject NP.

There were no remaining significant effects in the reading time measures.

3. General Discussion

The experiment reported above showed that the parser actively projects the syntactic position and grammatical features associated with a potential antecedent of the reciprocal before the antecedent has been reached in the input. Evidence for this comes not only from fixation times, but also from skipping rates. The experiment showed that readers had a tendency to skip over a two word noun phrase when the morphology on the head noun matched the predicted number of the phrase. These results have implications for theories of human parsing as well as theories of eye-movement control. We will consider each of these sets of implications in turn below.

3.1 Implications for theories of human parsing

The results reported here are consistent with a model in which the antecedent of a cataphoric dependency can be predicted, and assigned syntactic features, ahead of time, as proposed by Kazanina et al. (2007) and Kreiner et al. (2008). However, unlike those earlier studies, the present experiment established the relevant congruency effect in skipping rates, at a position *before* the critical word had been fixated in the input, thus lending further evidence for prediction. In the following paragraphs, we consider the parsing processes that might underlie the observed effects on skipping rates and

fixation durations that we observed. These results have implications for parsing theories because they clearly show that early sentence initial information is used for projecting upcoming structure.

In the introduction to the experimental section above, we assumed that the prediction of the plural NP leads to the computation of plural features in advance. This then leads to a facilitation of the parafoveal processing of the plural NP, resulting in the observed skipping rate differences. If this is so, one important question is what type of processing actually takes place in the parafovea. One possibility is that lexical access occurs for both the determiner *the* and the head noun *girl(s)* while the reader is fixating to the left of this region. Both words of this two-word phrase are accessed, and integrated with the predicted syntactic information. A second possibility appeals to the notion of shallow processing. According to this account, the phrase is only partially processed parafoveally. We assume that the determiner *the* is processed fully, as it is a short, high frequency function word. However, one possibility is that the head noun *girl(s)* is processed via a heuristic which involves checking the “s” plural marker, without necessarily going through a full lexical access process. According to this account, the parser predicts the plurality of the subject noun phrase, and also guesses the internal structure, assuming a highly probable determiner-noun combination. The parser predicts that the head noun should be marked for plurality, which in English usually means that an “s” morpheme is expected at the end of the word.⁴ According to this account, the skip-trials include cases where the plural prediction is satisfied through the parafoveal processing of the “s” morpheme at the end of the head noun, without the parser accessing the content of the noun. Thus, the reader will end up with only a partial interpretation of the phrase, because the content of the head noun will not have been retrieved, but will perceive the sentence as grammatical, because the morphology matches the requirements for the number feature. There are previous experiments that show that lexical decisions are facilitated if the syntactic category of the word is predictable from previous sentence context (e.g. Wright and Garrett (1984)). This result is consistent with a syntactic early predictive mechanism that is operative before full lexical access.

Based on the current data, it is not straightforward to distinguish between these two possibilities, but both are consistent with a predictive processing mechanism. In the second account, if lexical access has not yet occurred when the skip takes place, then this is by definition a case where behavior associated with prior prediction occurs before full bottom-up input has taken place, similarly to the case of DeLong et al. (2005) and Van Berkum et al. (2005). In the first account, where lexical access is assumed to take place even in the presence of a skip, one can argue for prior prediction on the basis of the extreme earliness of the effect, following a similar argument to that of Lau et al. (2006). After all, this is a case where information on the critical word is being processed from a point that is at least two words to the right of a fixation. It is hard to imagine a congruence effect occurring so early without prior prediction.

Of course the results we obtained also could be explained by non-predictive mechanisms, like the integration of the first subject NP with the sentence initial unbounded dependency. However the fact that we observed very early facilitation of processing of plural nouns in a predictive context makes this possibility unlikely. The results of the experiment imply that the parser processes at least the morphology of the parafoveal word if it is indicated by previous context that there should be a plural noun.

4. Of course, although this is a highly likely possibility, it is not the only one. For example, the head noun might not be morphologically marked for plurality (e.g. *the couple*, *the boy and the girl*, etc, or the head noun might not occur adjacent to the determiner (e.g. *the shop assistants*).

The results suggest that the parser not only expects a plural subject NP, but also that it projects the place where this antecedent of *each other* is anticipated (at the first possible subject NP position). In this respect these results are consistent with other experiments (Staub and Clifton, 2006), which suggest that the parser predicts the arrival of a particular structure if it is indicated by previous sentence context. For example Staub and Clifton (2006) showed that presence of *either* facilitated the fixation durations on phrases that start with *or*. Similarly in our experiment the parser using information from the dative reciprocal, actively predicted the site and the morphology of the upcoming subject NP.

In Staub and Clifton's study *either* induces a very high phrasal expectancy for the word *or*. In our case the reciprocal restricts the possibilities but does not direct the reader to a specific word. However along with the possible site of the antecedent noun phrase readers also predicted specific morphological information.

As we mentioned in the introduction, the sentence initial reciprocal is unbounded and it does not have to have its subject at the first position. However without waiting for bottom-up information confirming the place of the dependency site, the parser tries to construct the dependency at the first possible site, thus indicating an active, or eager search strategy (Kazanina et al., 2007). Moreover, results from skipping proportions and fixation times on first subject position indicate that the reader forms the dependency between the unbounded NP and its subject before encountering any verb information. Preverbal processing of dependencies has been shown in head final languages (Aoshima et al., 2004; Kamide et al., 2003). However the current experiment shows early processing of dependencies without verb information in English where generally the verb comes early in the sentence. This generalizes the previous evidence against head-driven parsing.

The early prediction of morphological information is consistent with recent results from MEG studies reported by Dikker et al. (2009) and Dikker et al. (2010).

In their MEG study, Dikker et al. (2009) showed a very early response for the nouns (around 130 ms) if there is a mismatch between the expected syntactic category and the category marking closed class morphemes. Dikker et al. (2010) showed that the parser not just reacts to closed class morphemes, but also to the form typicality of the words. In this study they compared nouns, with typical closed class category marking morphemes (like *farmer*) with nouns with typical properties of nouns, (like *movie*) and with neutral nouns, whose form does not have a clear bias towards either being a noun or a verb. They found a very early sensitivity to typical nouns and nouns with closed class morphemes, around 100 ms. These very early effects are hard to explain by full lexical processing of the words. The authors claim that their results not only indicate that the system predicts words categories, but that it also predicts the form features that are associated with that category. The skipping patterns obtained in our experiment are also in line with the results obtained by Dikker et al. As we mentioned earlier, it might be easier for the system to pick up the cues for plurality because sentence initial reciprocal highly constrains the features of the subject noun phrase.

The first pass times on the first subject region, showed that readers slowed down in the mismatch condition, indicating that the parser attempts to form the dependency between the sentence initial reciprocal and its possible antecedent at the first possible position, and this effect was replicated in the skipping probabilities. At the second subject position, there was a similar interaction effect in regression path time. This could indicate that readers attempt once more to form the dependency, this time with the second subject NP, in conditions where the initial attempt with the first subject has failed, due to a mismatch of features. If so, the increased reading times in the mismatch condition

could indicate the cost of retrieving sentence initial reciprocal and and integrate it with possible antecedent.

3.2 Implications for models of eye-movement control

There are three aspects of our results that have consequences for models of eye-movement control in reading. First, the results show that morphological information from the head noun must have been processed parafoveally, and second, this parafoveal information was accessed from a point at least two words to the right of the current fixation; in other words from word $N+2$. Finally, the results imply that syntactic prediction can influence saccade target location. All of these points have implications for models of eye-movement control in reading, and we will discuss them separately. In discussing models of eye-movement control, we will concentrate on two models, one of which, SWIFT, allows parallel processing of multiple words around the fixation (Kliegl et al., 2006), and the other, E-Z Reader allows only serial processing of words (Reichle et al., 2003). Of course, these are not the only two models of eye-movement control in the literature. However, these models are useful because they make clear and testable predictions, and because they represent very different approaches to the phenomena in question.

Current models of eye-movement control do not have specific mechanisms that allow for morphological pre-processing in the parafovea, and indeed, experimental work in eye-movement control has in general failed to show evidence that morphological information is accessed from a parafoveal word (Inhoff, 1989; Kambe, 2004; Lima, 1987). These earlier studies used display change techniques in which a word changes dynamically, as the reader's eye crosses an imaginary boundary in the text. The studies manipulated whether a morpheme from word $N+1$ was available for preview in parafoveal vision while the reader was fixating word N . The studies failed to find a facilitatory effect of morphological preview for fixations on word $N+1$. In other words, word $N+1$ was not read any faster when one of its morphemes had been previously available for preview, relative to when a non-morpheme string had been available for preview (see also Hyönä et al. (2004), Bertram and Hyönä (2007) and Hyönä and Pollatsek (1998) for studies failing to find parafoveal effects of morphology in reading Finnish compound nouns). However, none of these studies considered cases where this morphological information was highly predictable from the preceding syntactic context. In contrast, the experiment reported in the present paper not only manipulated the predictability of the morpheme, but also included conditions where the morphology both matched and mismatched the prediction. It might therefore be the case that morphological information is processed parafoveally only when this information is highly predictable from the context, as it was in our experiment reported here, and both match and mismatch conditions might be required in order to observe the effect.

Evidence consistent with our findings comes from a study in Hebrew reported by Deutsch et al. (2005), who showed that morphological information can be extracted parafoveally in Hebrew in cases where the parafoveal word has a common verbal pattern with the target word. Deutsch et al. (2005) also showed that if the parafoveal preview word was syntactically incongruent with the target word, then an inhibition effect in processing was observed. Interestingly, semantically biasing the context did not lead to parafoveal processing of the root morphemes. These results are important because they show processing of morphological information in the parafovea and these effects arise from syntactic context, but not when semantic information was manipulated. In a similar way in our experiment the sentence initial reciprocal constrains the upcoming information in several respects.

After reading the reciprocal, the parser expects to find the antecedent of the initial reciprocal phrase at the first possible position, which is immediately after the complimentizer *that*. In addition, the parser expects to have a plural and an animate (probably a NP related to a person) phrase. All of these constraints are possibly projected after reading the reciprocal, leading to facilitation of early processing of the projected noun, before fixating it.

An eye-tracking study by Underwood et al. (1990) showed that readers tend to look at the informative part of the word. If high information content is at the beginning of the word, as in the case of *engagement*, readers tend to look at the beginning of the word, compared to words like *underneath*. They claim that this effect could only be explained if the morphological information is processed parafoveally.

There is some evidence for processing of semantic (Yan et al., 2009) and to some extent morphological (Yen et al., 2008) information parafoveally in Chinese. For example Yen et al. (2008) showed either real words or pseudo words in the parafovea, and all words had the same initial morpheme with the target word. They showed that there were shorter fixations for the real words that have the same initial morpheme with the target word compared to pseudo words.

The second aspect of our results that has implications for models of eye-movement control is the fact that the relevant parafoveal information was extracted from word N+2. At first sight, this appears to be more compatible with models in which multiple words around the fixation can be processed in parallel, such as the SWIFT model (Kliegl et al., 2006), rather than models where word identification occurs serially, as in the E-Z reader model (Reichle et al., 2003). However, in the E-Z reader model, although words are processed serially, attention can move to word N+1 while word N is being fixated. If word N+1 is identified early enough, then the planned saccade to N+1 can be cancelled, and attention re-allocated to N+2. Thus, it is still technically possible in the E-Z reader model for processing of word N+2 to begin while word N is being fixated. In practice, within the E-Z reader model, this happens only under very specific circumstances, for example, when word N+1 is a very short or high frequency word. In our case, N+1 is the very high frequency word *the*, so it is possible that some aspects of the head noun could have been processed during the fixations before the article, even in the E-Z reader model.

In fact, however, evidence for parafoveal processing of word N+2 is very limited in the eye-movement literature. Using a display change technique, Angele et al. (2008) and Rayner et al. (2007) failed to find any evidence for preview benefit for N+2 (i.e. parafoveal viewing of N+2 did not lead to faster fixations when N+2 was subsequently fixated). However as Rayner et al. (2007) mention, in both of these experiments both N+1 and N+2 were relatively long words, apart from the second experiment of Rayner et al. (2007), where N+1 and N+2 were 3 or 4 letters long. In contrast, Kliegl et al. (2007) again in a display change experiment, found evidence for parafoveal processing of N+2, when the intervening word was 3 letters long. However the effect was observed in fixation durations on words N+1 and N. Also Risse et al. (2008) found a frequency effect on single fixation durations only when N+1 was 3 or 2 letters long and N+2 was not longer than 4 letters. In our experiment, where the nouns were preceded by three letter article, *the*, there may well have been some parafoveal processing of the noun following the article. Such processing may be facilitated if the article and noun are treated as a “word group” (Radach, 1996) for the purpose of eye-movement control. Radach (1996) showed that initial fixations on words that are preceded by a three letter word showed a single normal distribution over the whole combined two word string, consistent with the grouping of the two words together into a single visual object. Drieghe et al. (2008) found that these combined normal distributions are limited to situations where the three-letter

word preceding the noun is an article, although they argue that the apparent normal distribution may be better described as bimodal. However, in our stimuli, the critical noun was always preceded by the determiner *the*, so, if the *word grouping* hypothesis is correct, the two words were treated as a single unit, increasing the effectiveness of parafoveal preview of N+2. Indeed, even if readers were not grouping the article and the noun together, it is still possible that having a highly constraining syntactic context in addition to obvious parafoveal cues might have facilitated processing of the noun.

The final aspect of our results that has consequences for models of eye-movement control is the fact that skipping rates were affected by syntactic predictability. There is evidence in the literature that words that are predictable from context are skipped more often than those that are not (Ehrlich and Rayner, 1981; Rayner and Well, 1996; White et al., 2005). This fact can be accounted for in SWIFT (Kliegl et al., 2006), but also in E-Z Reader (Reichle et al., 2003), as mentioned in the introduction to the present paper. In SWIFT lexical activation for words builds up gradually, and starts to decline after it reaches the maximum. Saccades are directed towards the word with highest activation. If lexical processing is facilitated on word N+1 then the excitability starts to decline before the saccade program can be finalized. If the activation level for N+2 exceeds N+1, then the saccade is directed to N+2, resulting in a skip. In SWIFT the word can be skipped even if it is not fully lexically processed. However this explanation still does not account for the skipping rates in the present experiment. In our study, we manipulated not the predictability of a specific word, but the predictability of an abstract feature of the word, namely its plurality. Neither E-Z reader nor SWIFT include mechanisms that allow for this more abstract type of predictability to influence skipping rates.

4. Conclusion

Overall, the results of the experiment showed a strong case for predictive mechanisms in the formation of long distance dependencies, using eye movement measures. Also the results indicate a strong facilitation of parafoveal processing of words in a syntactically biased context. Clearly, further experiments are needed to follow up these findings, and to examine in more detail what types of information are extracted from the parafovea and under what types of conditions. However the experiment presented here addresses important and novel issues about the nature of the mechanisms and the time course of the effects of syntactic prediction.

References

- G. Altmann and Y. Kamide. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73:247–264, 1999.
- B. Angele, T.J. Slattery, J. Yang, R. Kliegl, and K. Rayner. Parafoveal processing in reading: Manipulating n+ 1 and n+ 2 previews simultaneously. *Visual cognition*, 16(6):697–707, 2008.
- S. Aoshima, C. Phillips, and A. Weinberg. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51:23–54, 2004.
- R. H. Baayen, D. J. Davidson, and D. M. Bates. Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, in press.

- D. A. Balota, A. Pollatsek, and K. Rayner. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17:364–390, 1985.
- D. Bates and D. Sarkar. *lme4: Linear mixed-effects models using S4 classes*, 2007. R package version 0.9975-11.
- R. Bertram and J. Hyönä. The interplay between parafoveal preview and morphological processing in reading. *Eye movements: A window on mind and brain*, pages 391–407, 2007.
- M. Brysbaert, D. Drieghe, and F. Vitu. Word skipping: Implications for theories of eye movement control in reading. In G. Underwood, editor, *Cognitive processes in eye guidance*, pages 53–77. Oxford University Press, Oxford, 2005.
- N. Chomsky. *Lectures on Government and Binding*. Foris, 1981.
- K. DeLong, T. Urbach, and M. Kutas. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117–1121, 2005.
- A. Deutsch, R. Frost, A. Pollatsek, and K. Rayner. Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Current Issues in Morphological Processing*, page 341, 2005.
- S. Dikker, H. Rabagliati, and L. Pykkänen. Sensitivity to syntax in visual cortex. *Cognition*, 110(3):293–321, 2009.
- S. Dikker, H. Rabagliati, T.A. Farmer, and L. Pykkänen. Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, 21(5):629, 2010.
- D. Drieghe, A. Pollatsek, A. Staub, and K. Rayner. The word grouping hypothesis and eye movements during reading. *Journal of experimental psychology. Learning, memory, and cognition*, 34(6):1552, 2008.
- S.F. Ehrlich and K. Rayner. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6):641–655, 1981.
- J. Hyönä and A. Pollatsek. Reading finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6):1612–1627, 1998.
- J. Hyönä, R. Bertram, and A. Pollatsek. Are long compound words identified serially via their constituents? evidence from an eye-movement-contingent display change study. *Memory and cognition*, 32(4):523, 2004.
- A.W. Inhoff. Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):544–555, 1989.
- T. F. Jaeger. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59:434–446, 2008.

- G. Kambe. Parafoveal processing of prefixed words during eye fixations in reading: Evidence against morphological influences on parafoveal preprocessing. *Perception and Psychophysics*, 66(2):279–292, 2004.
- Y. Kamide, G. T. M. Altmann, and S. L. Haywood. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49:133–156, 2003.
- N. Kazanina, E. Lau, M. Lieberman, M. Yoshida, and C. Phillips. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56:384–409, 2007.
- R. Kliegl, A. Nuthmann, and R. Engbert. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1):13–35, 2006.
- R. Kliegl, S. Risse, and J. Laubrock. Preview benefit and parafoveal-on-foveal effects from word n 2. *Journal of Experimental Psychology*, 33(5):1250–1255, 2007.
- H. Kreiner, P. Sturt, and S. Garrod. Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58:239–261, 2008.
- E. Lau, C. Stroud, S. Plesch, and C. Phillips. The role of prediction in rapid syntactic analysis. *Brain and Language*, 98:74–88, 2006.
- M-W Lee. Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33:51–73, 2004.
- R. Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, 2008.
- S. D. Lima. Morphological analysis in sentence reading. *Journal of Memory and Language*, 26(1): 84–99, 1987.
- K. McRae, M. J. Spivey-Knowlton, and M. K. Tanenhaus. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38:283–312, 1998.
- I. Mulders. *Transparent parsing: Head-driven processing of verb-final structures*. LOT, Utrecht, 2002.
- M. Pickering and S. Garrod. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11:105–110, 2007.
- B. L. Pritchett. Head position and parsing ambiguity. *Journal of Psycholinguistic Research*, 20(3): 251–270, 1991.
- R. Radach. *Blickbewegungen beim Lesen: Psychologische Aspekte der Determination von Fixationspositionen*. Waxmann Verlag, 1996.
- K. Rayner and A.D. Well. Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 1996.

- K. Rayner, J. Ashby, A. Pollatsek, and E. D. Reichle. 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(4):720–732, 2004.
- K. Rayner, B.J. Juhasz, and S.J. Brown. Do readers obtain preview benefit from word $n+2$? a test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1): 230–245, 2007.
- E. Reichle, K. Rayner, and A. Pollatsek. The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26:446–526, 2003.
- S. Risse, R. Engbert, and R. Kliegl. Eye-movement control in reading: Experimental and corpus-analysis challenges for a computational model. *Cognitive and cultural influences on eye movements*, pages 65–91, 2008.
- P. J. Schwanenflugel and E. J. Shoben. The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24:232–252, 1985.
- S. Sereno and K. Rayner. Measuring word recognition in reading: eye-movements and event-related potentials. *Trends in Cognitive Science*, 7:489–493, 2003.
- A. Staub and C. Clifton. Syntactic prediction in language comprehension: Evidence from either ... or. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32:425–436, 2006.
- M. J. Traxler and D. J. Foss. Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26:1266–1282, 2000.
- G. Underwood, S. Clews, and J. Everatt. How do readers know where to look next - local information distributions influence eye fixations. *Quarterly Journal of Experimental Psychology*, 42(1): 39–65, 1990.
- J. J. A. Van Berkum, C. M. Brown, P. Zwisterlood, V. Koojman, and P. Hagoort. Anticipating upcoming words in discourse: evidence from erps and reading times. *Journal of Experimental Psychology-Learning Memory and Cognition*, 31(3):443–467, 2005.
- R. P. G. Van Gompel and S. P. Liversedge. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology-Learning Memory and Cognition*, 29:128–139, 2003.
- S.J. White, K. Rayner, and S.P. Liversedge. The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic bulletin & review*, 12(3):466, 2005.
- B. Wright and M. Garrett. Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition*, 12(1):31–45, 1984.
- M. Yan, E. M. Richter, H. Shu, and R. Kliegl. Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, 16(3):561–566, 2009.

M. Yen, J. Tsai, O.J.L. Tzeng, and D. L. Hung. Eye movements and parafoveal word processing in reading Chinese. *Memory & Cognition*, 36(5):1033–1045, 2008.