The use of verb-specific information for prediction in sentence processing

Manabu Arai[a] and Frank Keller[b]

[a]JSPS, University of Tokyo, Department of Language and Information Sciences, Japan

[b]University of Edinburgh, School of Informatics, UK

Address correspondence to:

Manabu Arai

Department of Language and Information Sciences

University of Tokyo

3-8-1 Komaba, Meguro-ku

Tokyo, 153-8902

JAPAN

Tel: +81 03-5454-6786

Fax: +81 03-5454-6701

Email: m-arai@phiz.c.u-tokyo.ac.jp

Abstract

Recent research has shown that language comprehenders make predictions about upcoming

linguistic information. These studies demonstrate that the processor not only analyses the input

that it received but also predicts upcoming unseen elements. Two visual world experiments were

conducted to examine the type of syntactic information this prediction process has access to.

Experiment 1 examined whether the verb's subcategorization information is used for predicting a

direct object, by comparing transitive verbs (e.g., *punish*) to intransitive verbs (e.g., *disagree*).

Experiment 2 examined whether verb frequency information is used for predicting a reduced

relative clause by contrasting verbs that are infrequent in the past participle form (e.g., *watch*)

with ones that are frequent in that form (e.g., *record*). Both experiments showed that

comprehenders used lexically specific syntactic information to predict upcoming syntactic

structure; this information can be used to avoid garden paths in certain cases, as Experiment 2

demonstrated.

2

Introduction

In everyday conversation, we can often guess what a speaker is going to say next in the middle of their sentence. This illustrates our ability to predict upcoming linguistic material based on the information processed so far. This predictive behavior appears to be at work pervasively across linguistic domains and it is thus likely to be an integral component of human language processing.

In recent years, there have been a number of experimental studies that investigated predictive processes in sentence processing, across a range of experimental paradigms. For example, Altmann and Kamide (1999) reported a visual world study in which participants listened to sentences such as *the boy will eat/move the cake* while they viewed a scene containing a boy, a cake, and several distractor objects. When participants heard the verb *eat*, they tended to look at the cake more often compared to when they heard the verb *move*. This is because the semantics of the *eat* narrows possible upcoming referents down to one, but that of the *move* does not (there were several movable objects in the scene). This finding indicates that comprehenders are able to predict an upcoming direct object using the verb's selectional restrictions. Also, in a study by van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) using event-related brain potentials (ERPs), their Dutch participants first listened to a context that made a specific noun highly predictable. In the target sentence, they then heard an adjective that mismatched in gender with the predicted noun. Their results showed a differential ERP response on hearing the prediction-incongruent adjective compared to a no-context control condition. This finding indicates that comprehenders can predict lexical items based on context and access the grammatical features (such as gender) of the predicted items when they build syntactic

representations. A similar result was reported by Wicha, Moreno, and Kutas (2004), who observed a mismatch effect when an anticipated noun failed to match the preceding article in Spanish. DeLong, Urbach, and Kutas (2005) reported a comparable effect for English, using the distinction between the articles *a* and *an* instead of gender.

In fact, recent studies have revealed a great deal about exactly what kind of information listeners can use to predict upcoming information. Kamide, Scheepers, and Altmann (2003a) compared the processing of nominative-first constructions in German, such as *The hare-nom eats ... the cabbage-acc*, with accusative-first constructions, such as *The hare-acc eats ... the fox-nom,* and found that listeners were able to anticipate the correct postverbal argument once they had heard *eat,* i.e., they fixated the corresponding objects in a picture containing a hare, fox, cabbage, and distractor. This indicates that the processor is able to make predictions based on case-marking information in conjunction with the lexical semantics of the verb and the pre-verbal material. Kamide, Altmann and Haywood (2003b) reported similar results for English and Japanese, where Japanese is a particularly interesting case, as its verb-final syntax means that anticipation happens purely on the basis of the arguments of the verb and the case information they carry; the verb information itself becomes available only at the end of the sentence.

Prediction in sentence processing also has access to discourse information, as Kaiser and Trueswell (2004) have shown. In a study on Finnish, they compared subject-initial and object-initial structures. Object-initial structures are licensed in Finnish only if they introduce discourse-new material. Kaiser and Trueswell (2004) found that participants make anticipatory eye-movements to discourse-new material at the verb of object-initial structures, while subject-initial structures trigger no such predictions. This provides evidence that the processor is able to use the discourse context, in conjunction with word order information, to generate predictions about upcoming linguistic material. Another example of linguistic information that triggers prediction

is verb tense. Altmann and Kamide (2007) compared eye-movements for sentences such as *the man will drink ...* and *the man has drunk ...* in scenes depicting a full glass and an empty glass. They found more looks to the empty glass in the past tensed condition compared to the future tense condition; the inverse pattern was found for looks to the full glass, indicating that listeners make use of tense information in predicting upcoming linguistic material. Similarly, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) showed that prepositions can help anticipate upcoming material. They found that for sentences such as *put the cube inside the can*, on encountering *inside*, listeners immediately directed their attention to objects compatible with the semantic constraints of that preposition, thus providing another example of anticipation based on lexico-semantic information.

There is also evidence for prediction driven by syntactic constraints within a sentence. Staub and Clifton (2006) show that readers anticipate an *or* clause and thus process it faster if they have previously read an *either* clause, compared to a case where a preceding clause did not contain *either*. Similarly, studies that have shown syntactic parallelism in coordination (e.g., Frazier, Munn, & Clifton, 2000; Apel, Knoeferle, & Crocker, 2007; Sturt, Keller, & Dubey, 2010) can be interpreted as evidence for prediction: the processor anticipates that the second conjunct of a coordinate structure will have the same internal structure as the first conjunct (i.e., will be parallel to it), and a slowdown in the second conjunct can be observed if this prediction turns out to be incorrect.

As seen above, mounting evidence shows that the human sentence processor not only analyses the input that it has already received at any given point in the sentence, but also predicts upcoming material using syntactic, semantic, and discourse information. The current study contributes to this literature by examining another source of information, i.e., lexically specific syntactic information, as a cue for prediction.

It is well-known that verb-specific information is used by the sentence processor, though a number of authors have argued that it does not affect the initial analysis (e.g., Ferreira & Henderson, 1990; Frazier, 1987; Kennison, 2001; Pickering & Traxler, 2003; Pickering, Traxler, & Crocker, 2000; but see Mitchell, 1987). Other studies have found evidence for the immediate use of lexically specific information (e.g., Clifton, Frazier, & Connine, 1984; Stowe, Tanenhaus, & Carlson, 1991; Trueswell & Kim, 1998; Trueswell, Tanenhaus, & Kello, 1993; Garnsey, Pearlmutter, Myers, & Lotocky, 1997). This issue directly relates to the granularity of the information the processor has access to for generating predictions. Most previous studies measured processing difficulty at a disambiguating word or phrase *after* the target word, such as the disambiguating *by*-phrase following a structurally ambiguous verb (e.g., Trueswell, 1996). Hence it is still unclear whether verb-specific information is used immediately at the target word itself, or whether such information simply facilitates the integration of the disambiguating word into the current structure. In order to distinguish integration from prediction, we need to demonstrate the effect of lexically specific syntactic information at the target word itself; this rules out an alternative explanation in terms of integration.

Two previous studies suggest that verb specific syntactic information could drive a prediction. Firstly, Arai, van Gompel, and Scheepers (2007) found the effect of syntactic priming at the verb as a prediction about the upcoming word in the predicted structure. Their study suggests that listeners can immediately use verb's subcategorization information about the ditransitive structures (either a prepositional object or double object structure) when it has been primed from the previous sentence although finding the evidence only when the same verb was used between the prime and target sentences makes it impossible to rule out the possibility that the prediction was at least partly driven by the conceptual representations associated with individual verbs (but see Tooley & Traxler, 2009, who showed evidence against such an

explanation). Secondly, Boland (2005) looked at the role of argument structure information in prediction, comparing sentences with adjunct continuations like *slept for a while on the…* with sentences with argument continuations like *introduced her graciously to his…* The results showed that there were more looks to the argument entities than to the adjunct entities, suggesting that listeners can use argument status to inform their predictions about upcoming postverbal material. In the current study (Experiment 1), we compare transitive and intransitive verbs with respect to their effect on prediction, therefore extending Boland's results on argument prediction. We will return to a detailed discussion of Boland's experiments in the Discussion section of Experiment 1.

Models of prediction

Traditionally, models of human sentence processing have focused on capturing processing difficulty stemming from disambiguation and complexity. Model such as Gibson's (1998) Dependency Locality Theory (DLT) invoke integration cost and memory cost as explanations for processing complexity, but foresee only a limited role for prediction. In DLT, predicted syntactic structures need to be kept in working memory, which can limit the amount of memory available for the integration of new syntactic dependencies. Prediction can thus exacerbate integration cost effects, but does not in itself lead to processing difficulty.

Other theories have focused on disambiguation and modeled processing difficulty as the consequence of selecting the incorrect analysis initially (Frazier, 1987), of pruning low-probability analyses (Jurafsky, 1996), or of the competition between disambiguation constraints (McRae, Spivey-Knowlton, & Tanenhaus, 1998). These theories are not designed to model prediction directly; rather, they focus on disambiguation, which is modeled as the choice between a fixed set of alternatives (which can become more or less preferred as more input is processed).

7

There is no attempt to make fine-grained predictions by anticipating which syntactic rule or tree fragment is required to process the next word in the input stream.

Recent approaches to modeling sentence processing, however, have given prediction a prominent role. The most influential such theory is Hale's (2001) surprisal model, which accounts for processing difficulty based on the change in the probability distribution over possible analyses from one word to the next. (This can be formalized as the relative entropy of the two probability distributions, Levy, 2008). Words that trigger a small change in the probability distribution (i.e., that are predictable) are easy to process, while words that trigger a large change in the distribution are difficult to process.

Surprisal theory does not fix the linguistic representations over which probability distributions are computed and has been fleshed out in various ways in the literature. In Hale's (2001) original formulation, a probabilistic context-free grammar is used to compute the difference between $P(T_n)$ and $P(T_{n-1})$, where $P(T_n)$ is the distribution over syntactic structure at word w_n, the n-th word in the sentence. Later approaches (Hale, 2006) have used minimalist grammars to compute entropy reduction (conceptually related to surprisal), while other authors have opted for richer syntactic formalisms such as Tree Adjoining Grammar, which encodes specific structural assumptions (Demberg & Keller, 2009). All these approaches assume that prediction is fundamentally about syntactic structures (rather than just about words).

However, surprisal does not necessarily require recourse to syntactic structures. It can be formulated based on the difference between $P(w_1 ... w_n)$ and $P(w_1 ... w_{n-1})$, i.e., the probability distribution over word strings at word w_n and at the previous word in the sentence. In this case no grammar is required, an n-gram language model is sufficient. This approach is advocated by Frank (2009), who uses data from eye-tracking corpora to provide evidence for a purely word-based view of surprisal. Frank (2009) builds on the fact that a similar notion of word prediction

underlies Elman's (1991) simple recurrent network model of language, which can be adapted to compute surprisal estimates.

The current study reports two visual world experiments which examined the role of verb-specific information in prediction: subcategorization information (transitive or intransitive, Experiment 1) and verb frequency information (past participle vs. main verb frequency, Experiment 2). The aim is to determine whether these kinds of verb-specific information is accessed immediately during sentence processing, which is expected under a syntax-based model of prediction, but not by a purely word-based model such as Frank's (2009).

Experiment 1

Introduction

Experiment 1 used the visual world paradigm to investigate whether verb-specific syntactic information (about subcategorization) is used by the sentence processor to make predictions about an upcoming syntactic structure. We contrasted verbs that are almost exclusively used in the transitive structure such as *punished* in (1a) with ones that are almost exclusively used in the intransitive structure such as *disagreed* in (1b, c).

1a. Surprisingly, the nun punished the artist.
1b. Surprisingly, the nun disagreed with the artist.
1c. Surprisingly, the nun disagreed and the artist threw the kettle.

The intransitive verb was followed either by a preposition such as *with* (as in (1b)) or the conjunction *and* (as in (1c)). This allows us to examine whether comprehenders make predictions

about a prepositional complement as soon as a preposition is encountered. A sentence-initial adverbial phrase was included in an effort to add some context to the event of sentences.

While listening to these spoken sentences, participants saw a visual scene which depicted three entities: one animate entity that corresponded to subject (subject entity, henceforth), another animate entity that corresponded to direct object for the transitive verb condition but as a prepositional complement or as a continuation following the conjunction for the intransitive verb condition (target entity), and an inanimate entity that was not mentioned and always implausible as a direct object of the transitive verb (distractor entity).

Figure 1. Example stimulus for Experiment 1.

If verb-specific subcategorization information can immediately be accessed and used at the verb, then we expect participants to predict a postverbal direct object when they hear a transitive verb, but not when they hear an intransitive verb. Therefore, we expect participants to make anticipatory eye-movements toward the target entity that corresponds to the predicted direct object (i.e., patient argument, *artist* in the example above) on hearing a transitive verb but not an intransitive verb. Specifically, we predict that participants would look more at the target entity

with a transitive verb compared to an intransitive verb. Also, looks to the subject entity are expected to decrease over time with the former type of verbs compared to the latter. On the contrary, if people do not make predictions using subcategorization information and only have access to the syntactic category of a verb, participants initially extract identical information from the two types of verbs. Thus, there should not be any difference in their eye-movements depending on whether they hear a transitive or intransitive verb.

Furthermore, we predict that participants would make anticipatory eye-movements toward the same entity (i.e., *artist*) on hearing a preposition following an intransitive verb because the target entity, but not the distractor, is plausible as a complement of the prepositional phrase being constructed. It has been shown that people make such anticipatory eye-movements with prepositions (e.g., Chambers et al., 2002). On the contrary, these anticipatory eye-movements should not occur on hearing the conjunction *and* following an intransitive verb as almost any type of continuation is possible (e.g., …*and kicked the kettle*). Therefore, participants should look more at the target entity on hearing the preposition compared to when hearing the conjunction *and*.

Method

Participants

Thirty-three participants from the University of Edinburgh student community took part in the experiment. All were native speakers of English with normal visual acuity. They received £5 in exchange for their participation. Three participants were excluded due to recording difficulties.

Materials

Twenty-four items were constructed, each consisting of a pre-recorded spoken sentence and a semi-realistic visual scene. Twenty-four transitive and intransitive verbs were used and there was no repetition of verbs across items. By consulting the COMLEX syntactic dictionary (Grishman, Macleod, & Meyers, 1994), we selected transitive verbs that do not have an intransitive entry and intransitive verbs that either do not have a transitive entry (i.e., one not followed by a NP) or have a transitive entry but require a very specific type of NPs (e.g., *he screamed her name*). There are nine of the latter type of verbs in total (*shout, scream, wave, bow, laugh, wink, jump, whisper,* and *yell*) and none of those verbs can form a grammatical sentence by taking any of the depicted entities as its direct object (e.g., **The cheerleader screamed the runner/cheque*). The sentences were read by a Scottish male speaker with neutral accent and recorded in 16 kHz mono format. All the sentences were recorded individually, i.e., no cross-splicing was applied for the lexically identical verbs in the two intransitive conditions, although this may possibly result in subtle prosodic difference between the two verbs. This is because it is known that disruption of intrinsic prosodic information by cross-splicing could affect processing efficiency (e.g., Tyler & Warren, 1987). Table 1 shows the mean verb duration and the mean onset of preposition/conjunction, postverbal determiner, and postverbal noun following the verb onset. The duration of the verb differed numerically across the conditions, and an independent t-test showed that the verb duration in the two intransitive conditions was different from that in the Transitive condition although the difference was marginal for the Intransitive+PREP condition ($t(45.6) = 1.74, p = 0.09$ for Transitive vs. Intransitive+PREP; $t(45.6) = 3.27, p < 0.01$ for Transitive vs. Intransitive+CONJ). Interestingly, the difference in the verb duration between the two intransitive conditions, although lexically identical, was also significant by a paired t-test ($t(23) = 3.13, p < 0.01$).

1

Table 1. The mean verb duration and the mean onset of preposition/conjunction, postverbal determiner, and postverbal noun following the verb onset for each condition in milliseconds. Verb lengths in characters are given in brackets.

	Verb duration	Preposition/Conjunction	Determiner	Noun
Transitive	803 (8.0)		1310	1451
Intransitive+PREP	746 (6.7)	1242	2113	2268
Intransitive+CONJ	706 (6.7)	1253	2122	2307

The visual scenes were constructed by editing commercial art-clip images with graphic software and saved in 24-bit colour bitmap format (1024 x 768 pixels resolution).

The three entities in each picture were arranged in a triangular fashion (as can be seen in Figure 1). We counterbalanced the position of these entities (subject, target, and distractor) to avoid biasing viewer's scanning pattern. The pictures stayed the same across experimental conditions.

Design and procedure

Three versions of auditory sentences, which correspond to the three conditions described in (1), and one version of visual scenes were prepared for each experimental item. Additionally, 48 pictures with auditory sentences were included as fillers. None of the filler sentences used either intransitive or transitive verbs; all were copula-verb constructions. As with the experimental target items, the fillers also employed visual scenes with three entities in a triangular arrangement. Each participant saw only one of the three experimental conditions of each item and the same number of items in each condition. The materials were presented in a randomized order with the constraint that each experimental item was preceded by at least one filler item. Each experimental session started with two fillers.

Participants were seated at a distance of approximately 75 cm from a 21" color monitor running at 120 Hz refresh rate. Participants' eye-movements during each trial were recorded with SR Research EyeLink II head-mounted eye-tracking system at the sampling rate of 500 Hz and spatial accuracy of 0.5°. Viewing was binocular, but only the participant's dominant eye was tracked as determined by a simple parallax test prior to the experiment. Although not restricted in any way, participants were instructed to keep head movements to a minimum during the experiment. Auditory stimuli were presented via a desktop speaker. The eye-tracker continuously recorded onsets and offsets of fixations (as defined by acceleration and velocity thresholds) together with corresponding spatial coordinates (in pixels).

The experiment began with the adjustment of the infrared cameras of the tracker, a procedure that took about one minute per participant. A brief calibration procedure was next performed during which the participant had to look at a fixation cross in nine different positions on the screen. This procedure was repeated after a short break half way through the experiment and whenever measurement accuracy appeared insufficient. We informed participants that the spoken sentences would always refer to the visual scenes, and that it was important to pay attention both to the visual scene and the spoken sentence. Auditory sentence presentation always started after a 1000 ms preview period following the onset of the picture. The picture stayed on the screen for seven seconds, after which the next trial was initiated. To keep participants focused, six fillers were followed by a written prompt indicating that participants had to verbally describe the previous scene (participants typically responded by repeating the auditory sentence that they had just heard). Each experimental session took 30 to 40 minutes to complete.

Data analysis and results

The X-Y coordinates of individual fixations were classified into one of four segmented areas in the visual scene by using colour-coded picture templates: subject entity, target entity, distractor entity, and background. The areas of three visual entities were defined by expanding the contour of each entity by 30 pixels. The other area that did not belong to these areas of visual entities was coded as background. We also excluded extremely short fixations (with duration of less than 100 ms) as they usually do not reflect visual information processing (e.g., Rayner & Pollatsek, 1989). To analyse eye-movements with reference to corresponding spoken sentences, the onsets of the verb, the preposition/conjunction, the postverbal determiner, and the postverbal noun in each spoken sentence were manually marked in millisecond-resolution using sound editing software.

Of particular interest are the fixations that were made following the verb onset. If language comprehenders make use of verb-specific syntactic information to predict what structure the verb would appear in (i.e., intransitive or transitive), we expect to observe anticipatory eye-movements to the target entity as a plausible direct object entity on hearing a transitive verb such as *punish* but fewer such eye-movements on hearing an intransitive verb such as *disagree*. It is important to note that, in our items, only the target entity can plausibly be a direct object of a transitive verb (**the nun punished the kettle*). Furthermore, we also predict that participants would make anticipatory eye-movements toward the same entity on hearing a preposition following an intransitive verb but not on hearing the conjunction *and* following it. This is because the preposition in our items always takes the target entity as a plausible complement whereas taking the distractor entity results in a highly implausible situation although not entirely impossible (*?The nun disagreed with the kettle*). On the contrary, the conjunction *and* can be followed by any type of continuation and thus it is highly unlikely that participants could make a prediction about what would follow the conjunction.

1

We first analyze the number of trials in which participants made saccadic eye-movements toward the target entity while listening to the verb (i.e., for the duration of individual verbs). Saccades are the more appropriate measure than fixations here as it is possible that a saccade might have initiated before the verb onset (i.e., independently of the verb information) although the following fixation occurred after the verb onset. We also report the number of trials in which they were looking at the target entity at the onset of the postverbal material. Next, we report the analysis on the amount of visual attention to the target entity over a certain time interval after participants heard a verb and a postverbal word. More specifically, we conducted analyses on the logit of gazes following the verb onset and that following the postverbal word onset. This analysis has the advantage of showing how the looks to a particular entity change over time and can corroborate the finding from the saccade probability analysis. In both analyses, we used Linear Mixed Effects (LME) models (e.g., Baayen, Davidson, & Bates, 2008; Barr, 2008). LME models can simultaneously include random effects of participants and items, as well as experimental manipulations as fixed effects. Furthermore, they are robust in handling missing data points, which are fairly common in eye-movement measures and cause a problem for traditional procedures such as analysis of variance. For all analyses, we report coefficients, t-values, 95% confidence intervals, and significance levels for all fixed factors. The 95% confidence intervals are Highest Posterior Density intervals computed using Monte Carlo Markov chain sampling and the significance levels are based on the posterior distribution computed using the same sampling technique.

In all the analyses reported in this paper, we constructed optimal LME models of the data as follows. We started with a model that included all the main effects of the fixed factors, as well as the covariate Time Window, and all interactions between these factors. We then tested whether removing the highest order interaction reduced model fit significantly, as indicated by a χ^2 test on

the log-likelihood values of the models. The interaction was removed if this did not result in a significant reduction in model fit; this procedure was iterated for all interactions and main effects, until the minimal model that optimized model fit was found. No model selection was performed on the random component of the models; these always included the random factors participants and items. In order to reduce collinearity, all fixed factors were centered (unless stated otherwise).

The first analysis examined the saccades that were initiated from outside of the target entity and landed inside that entity. We calculated the proportion of trials in which participants launched at least one such saccade within the duration of the verb. Table 2 (first row) shows the mean percentages of such trials.

Table 2. The percentages of trials where participants launched at least one saccade toward the target entity within the duration of the verb and the percentages of trials where participants were looking at the target entity at the onset of the postverbal material.

Condition	Transitive	Intransitive+PREP	Intransitive+CONJ
Saccades after verb onset and before postverbal onset	50.0%	29.5%	27.0%
Fixations at postverbal onset	50.6%	42.7%	39.7%

An LME model was used to analyze the probability of trials where participants launched at least one saccade to the target entity within the verb duration using a binomial family function (Bates, 2010). This model included the random factors participants and items and one fixed factor with three levels: Transitive verb, Intransitive verb with a preposition (Intransitive+PREP), and Intransitive verb with a conjunction (Intransitive+CONJ). The transitive condition was treated as a baseline against which the effects of two Intransitive conditions are compared.

1

The results showed that there was a significant difference between the Transitive condition and the two intransitive conditions respectively (Transitive vs. Intransitive+PREP, coefficient = -0.57, z = 2.81, p < .01; Transitive vs. Intransitive+CONJ, coefficient = -0.70, z = 3.43, p < .001). The results suggest that the type of the verb, transitive or intransitive, affected the probability of launching saccadic eye-movements toward the target entity; participants were more likely to look at the target entity on hearing transitive verbs than intransitive verbs. We also conducted an analysis including only the two intransitive conditions. The results confirmed that there was no difference between the two conditions (coefficient = -0.15, z = 0.72, p = 0.47).

We also examined the probability of trials where participants were looking at the target entity at the very onset of the postverbal material (Table 2, second row). We used the same LME model and the results suggest that there was a difference between the transitive condition and the two intransitive conditions, although the difference between Transitive and Intransitive+PREP was marginal (Transitive vs. Intransitive+PREP, coefficient = -0.34, z = 1.84, p = .07; Transitive vs. Intransitive+CONJ, coefficient = -0.47, z = 2.51, p = .01). This may reflect some subtle prosodic difference with the verbs between the two intransitive conditions.

Gaze logit analysis

We now examine the fixations across time frames. This analysis allows us to examine moment-to-moment changes of participants' eye-movements over time as a function of specific linguistic information. We first converted the fixations into gazes by accumulating all the consecutive fixations on any depicted entity until another entity (or the background) was fixated. To examine the probability of gazes to the target entity following the verb onset, we excluded the trials on which participants were already looking at the target entity at the onset of the verb. Any difference between the conditions thus reflects the processing of the verb information but not of

1

anything before the verb. This removed 18% of the whole data (130 out of 720 trials) and a χ^2 test shows that the number of exclusions did not differ across the three conditions ($\chi^2 = 0.8$, $df = 2$, $p = 0.7$). Figure 2 plots the gaze probability to the target entity aggregated for each condition in each 20 ms time bin over 1500 ms, starting from the verb onset.
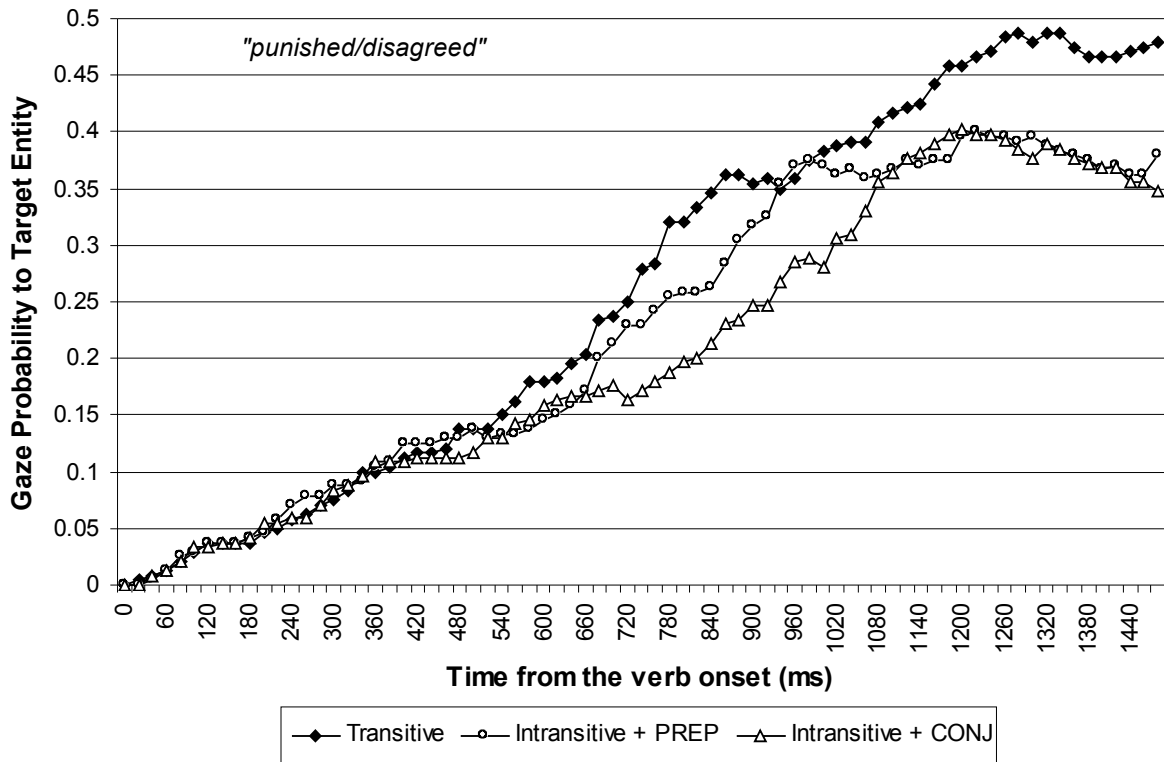

Figure 2. Gaze probability to the target entity aggregated for each condition in each 20 ms time bin over 1500 ms, starting from the verb onset. The two vertical lines mark the average verb offset and the average onset of the postverbal material..

To analyse the data, we calculated the empirical logit of the gaze probabilities for the target object for each 100 ms time bin over the 800 ms following the verb onset using the following function (e.g., Barr, 2008):

1

$$\eta' = \ln((y + 0.5)/(n - y + 0.5))$$

where n the total number of looks within each 100 ms time window to all the objects in a scene (including background). Unlike a probability scale, the logit scale can be infinitely positive or negative. The measure yields a score of zero if looks to a target entity (y) occur exactly half of the time compared to looks to all other entities, and yields a positive value if there are more looks to the target entity than looks to all other entities. As Table 1 shows, our stimuli contained a relatively large pause between the verb offset and the onset of the postverbal material and it is possible that the pause could have encouraged predictions of upcoming material to some extent. Thus, we restricted the interval for the analysis to the time windows up to 800 ms post onset, which approximately corresponds to the mean duration of the verb (752 ms, SD = 114). We also excluded the first 100 ms from the analysis given that the minimum duration in which perceived linguistic information can bias a fixation is thought to be between 100 – 200 ms.[i] Therefore, we analyzsed the 700 ms time interval (from 100 ms to 800 ms following the verb onset) where the effect of verb information itself on eye-movements, but not that of any postverbal material, is most likely to occur. We conducted our analysis using LME on the logit of gaze probabilitiesy to the patient target entity, including Verb Type (transitive vs. intransitive verbs, entered as three levels) as a predictor variable and seven 100 ms time windows as a covariate (Time Window) to take into account the change of fixation gaze probabilities over the time period (Barr, 2008). In the analysis below, the Transitive condition is again treated as the baseline against which the two intransitive conditions are compared.

The summary of the results from the optimal model, which was achieved following the same procedure as that explained earlier, is shown in Table 3. There was a significant effect of

Time Window. Participants tended to look increasingly at the patient target entity over the course of time. A main effect of Intransitive+CONJ was significant and that of Intransitive+PREP was marginal ($p = 0.08$), which suggests that participants made more looks toward the target entity on hearing a transitive verb than on hearing an intransitive verb, although the difference was not fully reliable when the intransitive verb was later followed by a preposition. Importantly, the interaction between Intransitive+CONJ and Time Window and that between Intransitive+PREP and Time Window were both significant, suggesting that there was a greater increase of fixations to the target entity over this time period with transitive verbs than with the intransitive verb.

Table 3. Analysis of gaze logits to the target entity for 800 ms interval from 100 ms to 800 ms following the verb onset for the three conditions.

Factor	coefficient	t	CI	sig.
(Intercept)	-1.25			
Intransitive+PREP	-0.10	1.77	0.11	
Intransitive+CONJ	-0.19	3.48	0.11	<.001
Time Window	0.19	9.55	0.04	<.001
Intransitive+PREP x Time Window	-0.06	2.24	0.05	<.05
Intransitive+CONJ x Time Window	-0.09	3.25	0.05	<.01

*p<0.05, **p<0.01, ***p<0.001

Visual inspection of Figure 2 indicates that there are more looks to the target entity in the Intransitive+PREP condition compared to the Intransitive+CONJ condition, which was not expected as the verbs in both conditions are lexically identical. To see if this difference is statistically supported, we conducted an additional analysis including only the two intransitive conditions using the same LME model, so that one of the intransitive conditions is now treated as a baseline and the two conditions are directly compared. The results from the optimal model showed a marginally significant main effect of the conditions (coefficient = -0.09, $t = 1.75$, $p =$

0.08). This suggests that there was a trend that participants looked at the target entity more when the intransitive verb was later followed by a preposition than when it was followed by the conjunction *and* although it was statistically not fully supported. This is possibly due to a subtle acoustic difference between the auditory stimuli in the two conditions. As seen earlier, the duration of the verb in the Intransitive+CONJ condition was significantly shorter than that in the Intransitive+PREP condition. Possibly, the latter verb contained prosodic information that somewhat signalled a continuation of a sentence, which may have led to marginally more looks to the target entity compared to the verb in the Intransitive+CONJ condition.[ii]

Our next analysis concerns the gazes to the other entities in the display, i.e., the subject and distractor entities. Such an analysis is important in order to establish that the effect we are observing is specific to the target entity. It is conceivable that the other entities may show the same difference between transitive and intransitive conditions, thus indicating that the effect is an artefact of our experimental materials, rather than a genuine effect of anticipatory eye-movements triggered by verb subcategorization information.
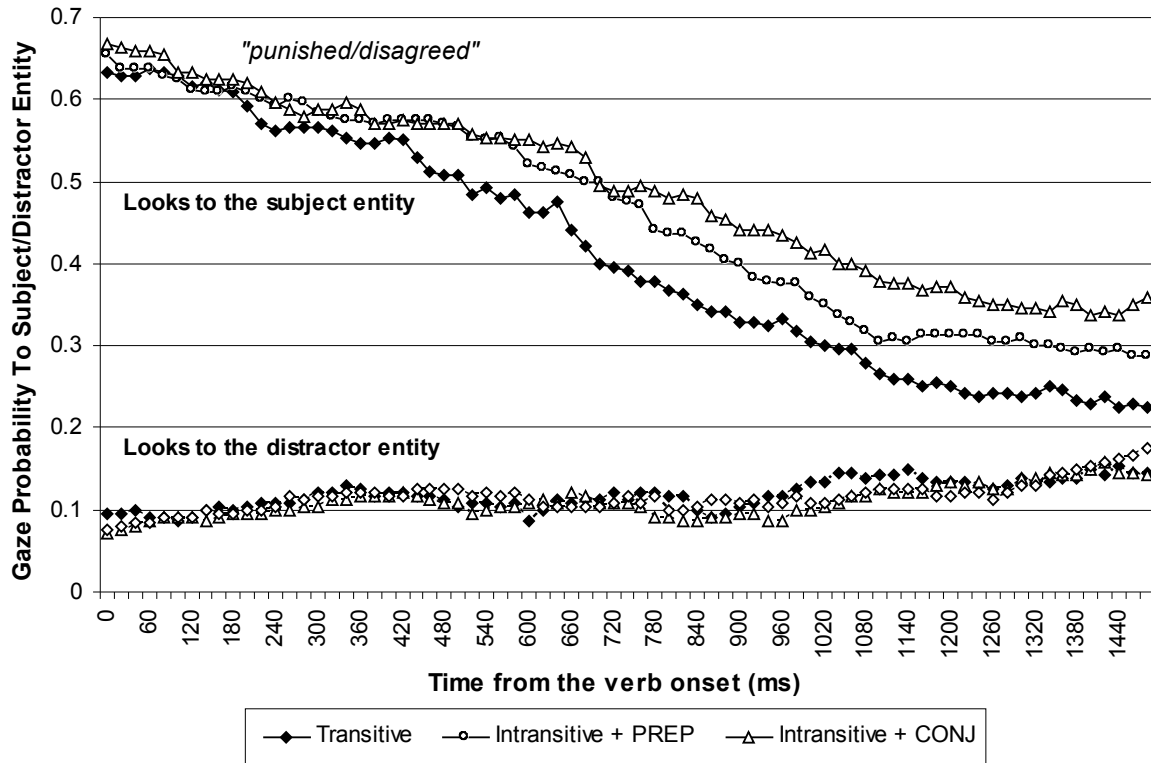
Figure 3. Gaze probability to the subject and distractor entities aggregated for each condition in each 20 ms time bin over 1500 ms, starting from the verb onset. The mean offset of the verb and the mean onset of the postverbal material are marked by two vertical lines.

Figure 3 shows the gaze probabilities to the subject and distractor entities. Visual inspection of the graph shows that the pattern at the subject entity is the opposite of the one at the target entity: the gaze probability is lowest in the transitive condition, intermediate in the intransitive+PREP condition, and highest in the intransitive+CONJ condition. This is not surprising, as it indicates that participants are fixating the subject entity when they are not fixating the target entity (the object). We conducted an LME analysis for the logit of the gaze probabilities to the subject entity for the interval between 100 ms and 800 ms following the verb onset. Again, Transitive condition was treated as a baseline in our model. The results from the optimal model showed the main effect of Intransitive+CONJ (coefficient $=0.19$, $t = 2.81$, $p <$

2

0.01) as well as the interaction between Intransitive+CONJ and Time Window (coefficient = 0.07, $t = 2.01$, $p < 0.05$). Also the main effect of Intransitive+PREP was significant (coefficient = 0.17, $t = 2.53$, $p < 0.05$), but the interaction between Intransitive+PREP and Time Window was not ($t < 2$). This indicates that gaze probability decreased more quickly in the Transitive condition compared to the Intransitive+CONJ condition while it did not when compared to the Intransitive+PREP condition; again this is the inverse of the findings at the target entity, ruling out that the result is due to an artefact. The probability of looking at the distractor entity was fairly low throughout and did not differ across the conditions, which was confirmed by conducting another LME analysis in the same way for the looks to the distractor entity ($t < 2$ for all the main effects and interaction).

Our next set of analyses concerned a different region of the sentence: we analysed the gazes following the preposition/conjunction onset. Again, in order to clearly see the effect of the postverbal elements on the eye-movements, we excluded trials where participants were already looking at the target entity at the onset of the postverbal elements. The number of these trials is quite high (41%, 197 out of 480 trials), though this is not surprising as the target entity already attracted a considerable amount of visual attention by the time participants heard the postverbal material (see Figure 2). A χ^2 test yet showed that the number of excluded trials did not differ between conditions ($\chi^2 = 0.25$, $df = 1$, $p = 0.62$). Figure 4 shows the probability of gazes to the target entity, which is a plausible continuation following a postverbal preposition (*artist* following *disagreed with*), aggregated for the two intransitive types in each 20 ms time bin over 1000 ms, starting from the preposition/conjunction onset.
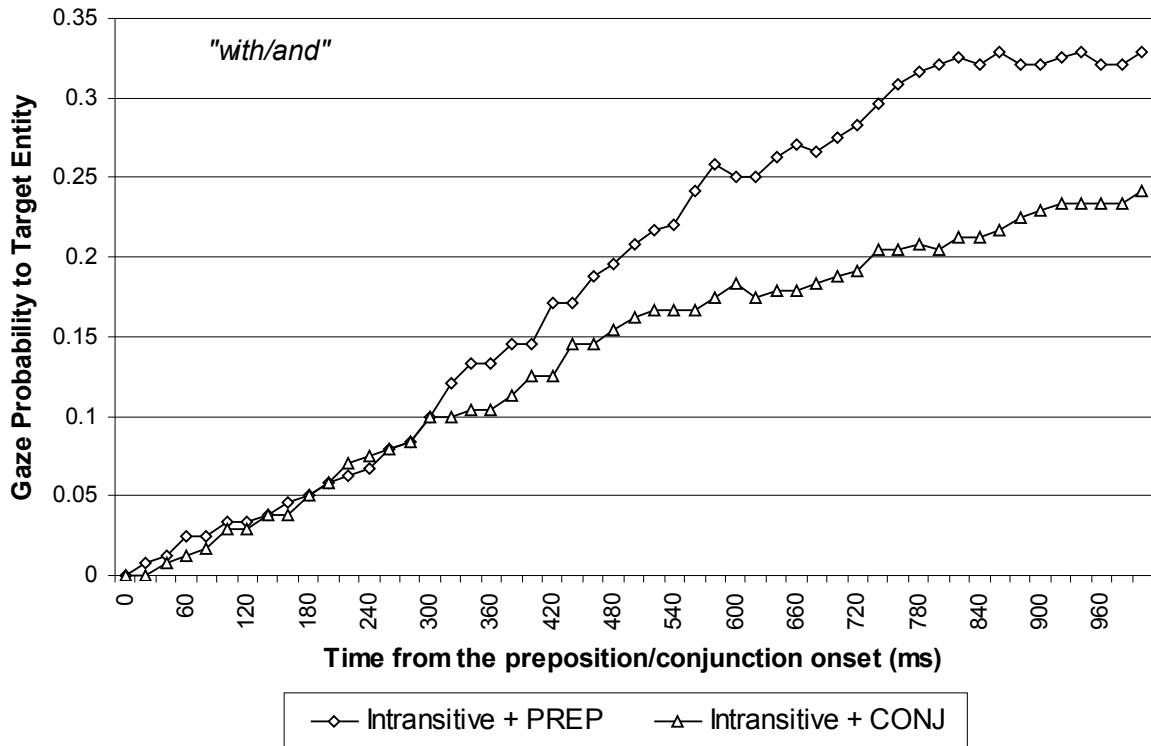
2

Figure 4. Gaze probability to the target entity aggregated for the two intransitive conditions in each 20 ms time bin over 1000 ms, starting from the preposition/conjunction onset. The average onset of the noun phrase following the preposition/conjunction is marked by a vertical line.

We conducted a LME analysis on the logit of gaze probabilities for the 700 ms interval from 100 ms to 800 ms following the preposition/conjunction onset for the two intransitive conditions. The 800 ms cut-off point was decided because the minimum onset of the determiner following the preposition/conjunction across items was 700 ms. We included Intransitive Type (Intransitive+PREP vs. Intransitive+CONJ) as a predictor variable and Time Window as a covariate in the model, along with two random factors (participants and items). The results from the model are summarized in Table 4.

2

Table 4. Analysis of gaze logits to the target entity for 700 ms interval from 100 to 800 ms following the preposition/conjunction onset for the two intransitive conditions.

Factor	coefficient	t	CI	sig.
(Intercept)	-0.66			
Intransitive+CONJ	-0.35	6.09	0.11	<.001
Time Window	0.16	7.94	0.04	<.001
Intransitive+CONJ x Time Window	-0.11	3.79	0.06	<.001

*p<0.05, **p<0.01, ***p<0.001

There was a main effect of Time Window, indicating that participants increased looks to the target entity over this time period. There also was a main effect of Intransitive Type, which shows that the participants looked more at the target entity during this time period on hearing a preposition than on hearing a conjunction. Most importantly, the interaction between Intransitive Type and Time Window was significant. This suggests that participants' looks to the target entity increased more over time when they heard the preposition than when they heard the conjunction and that the postverbal preposition has driven the anticipatory eye-movements toward the entity that is plausible as a complement of the predicted prepositional phrase.

Discussion

The results from the saccade probability analysis demonstrated that participants were more likely to look at the upcoming target entity with transitive verbs than with intransitive verbs. Furthermore, the analysis of gaze logits showed that looks to the target entity increased more over time with transitive verbs than with intransitive verbs. Importantly, these results were obtained for anticipatory eye-movements, i.e., eye-movements that were made before any linguistic information following the verb became available. Both analyses thus suggested that participants used verb-specific subcategorization information immediately at the verb to make a prediction about the syntactic structure that the verb licenses (either transitive or intransitive).

2

Furthermore, the analysis of gaze logits following the preposition/conjunction showed that participants made predictions about an upcoming complement as soon as they heard a preposition, but not when they heard a conjunction. This suggests that the dependency relationship between a preposition and a lexical head is immediately realized at the preposition and predictions about the upcoming complement are generated without delay.

The results of Experiment 1 therefore showed clear evidence that comprehenders access lexically-specific subcategorization information at the verb and use it to make predictions about an upcoming direct object. In other words, comprehenders construct a syntactic analysis at the verb using verb-specific subcategorization information. The results of Experiment 1 cannot be explained in terms of integration of subsequent information, as the effect was observed before any postverbal material was heard.

It is instructive to compare our Experiment 1 to the results of Boland (2005), who investigated some of the same issues. Recall that Boland tested sentences with adjunct continuations like *slept for a while on the…* and compared them to sentences with argument continuations like *introduced her graciously to his…*. She further manipulated the typicality of the adjunct/argument entities (e.g., *slept for a while on the <u>bed</u>* or <u>bus</u>). The results showed that there were more looks to the argument entities than to the adjunct entities. However, they also found a typicality effect in both the adjunct conditions and the argument condition (i.e., more looks to typical than to atypical entities), which suggests that participants predicted not only arguments but also adjuncts. The interpretation of Boland's results is complicated by differences with respect to where in the sentence the adjunct/argument can appear: the location and instrument entities cannot directly follow the verb (**slept the bed/*beat the stick*) while this was possible for the recipient entities in most of her items (e.g., *introduced his doctor*). Furthermore, Boland's analysis is based on looks to different pictures across conditions (e.g., bed and

teenager), which makes it impossible to separate the effect of the visual stimuli from that of the verbs themselves. One of Boland's experiments did use the same picture and compared verbs that take a recipient argument (e.g., *mentioned it right away to the owner*) with verbs that take a benefactive adjunct (e.g., *fixed it hurriedly for the owners*). Although an effect of verb type was again found, the interpretation of the results is confounded by the fact that the former verbs in Boland's materials can normally take the entity as a direct object (*mentioned the owner*), whereas the latter verbs cannot (*fixed the owners). Therefore, the effect found by Boland is arguably based on whether a depicted entity can be a part of the sentence, rather than on whether the verb itself can take a direct object (in fact both verbs *mention* and *fix* can take a direct object, just that no suitable entity for *fix* appeared in the scene).

While none of the problems with Boland's (2005) study was present in our experiment, thematic fit could be a potential confound for our results. It is known that thematic fit can influence the processing of the upcoming information, as the combination of subject noun and verb information imposes constraints (selectional restrictions) on possible syntactic analyses (e.g., McRae et al., 1998) or admissible lexical items (Kamide et al., 2003b). However, in the current experiment, we held the subject noun constant across conditions and included the verbs that are not usually associated with the subject noun (i.e., punishing or disagreeing is not the typical action expected for a nun). We can therefore be confident that we have minimized the potential confounding effect of thematic fit.

To summarize, Experiment 1 provided evidence that verb-specific syntactic information (viz., information about subcategorization frames) is available to the sentence processor for generating predictions. In Experiment 2, we will extend our study of prediction to another type of verb information that is potentially useful for making predictions, i.e., frequency information.

Specifically, we will investigate whether the frequency with which a verb occurs in a particular form plays a role in prediction.

Experiment 2

Introduction

The aim of Experiment 2 was to investigate the use of frequency information in prediction. More specifically, we examined whether the frequency of a given verb form (past particle vs. main verb) is used by the sentence processor to generate structural predictions. Previous research demonstrated that lexical frequency information plays a role in resolving syntactic ambiguities, but has not specifically looked at prediction. Trueswell (1996) tested two groups of verbs using the self-paced reading paradigm: verbs that are frequently used in the past participle such as *describe* in (2a) and those that are less frequently used in that form such as *scratch* in (2b).

2a. The necklace described by the lady was quite beautiful.
2b. The sofa scratched by the cat was badly damaged.

The part of the sentence up to the verb was syntactically ambiguous between a main clause analysis and a reduced relative clause analysis, although semantic information strongly biases toward the latter interpretation. Trueswell (1996) compared reading time for the reduced relative clause in (2) with the unreduced relative clause counterpart including *that was*. His results showed that when the relative clause analysis was supported by the lexical frequency information as well as by semantic information, processing difficulty due to syntactic misanalysis for sentences like (2b) was eliminated. This paper also showed that lexical frequency alone was not

enough to preempt syntactic misanalysis (Experiment 2). He thus argued that lexical frequency information, along with semantic support, boosted the availability of the relative clause structure and eliminated processing difficulty associated with the syntactic misanalysis. It is possible given Trueswell's results that lexically specific information may have helped to quickly resolve syntactic ambiguities on encountering the *by*-phrase, i.e., it helped integrating the disambiguating information, but it is not involved in predicting the correct reading already at the verb. Therefore, the current study attempts to see if verb form frequency information plays a role not only in the process of resolving syntactic ambiguities but also in the process of predicting a correct structure. Specifically, we examine whether the relative clause structure is activated using lexical frequency immediately at the verb, rather than at the disambiguating postverbal phrase. The visual world paradigm used in the present study is particularly suited for teasing apart these processes (i.e., prediction and integration), as it enables us to examine participants' syntactic choice through differential looks to entities in visual scenes. Since we analyse effects that occur before postverbal material is heard, the results should reflect the processing of verb information, but not of postverbal disambiguating information, and thus tell us about predictive processes at the verb.

We tested spoken sentences of the same construction as in Trueswell (1996) in four experimental conditions below, see (3, 4) for examples.

3a. The videotape watched by the student was found under the chair.

3b. The videotape that was watched by the student was found under the chair.

4a. The song recorded by the nun was about the flower.

4b. The song that was recorded by the nun was about the flower.

The sentence contained either a verb such as *watch* that is infrequently used as a past participle (relative to other forms, low PastP verb henceforth) or a verb such as *record* that is frequently used in that form (high PastP verb). Also, the verb appeared either in a reduced relative clause (3a, 4a) or in an unreduced relative clause (3b, 4b). Participants saw a visual scene paired with a spoken sentence. Figure 5 shows example scenes for (3, 4).
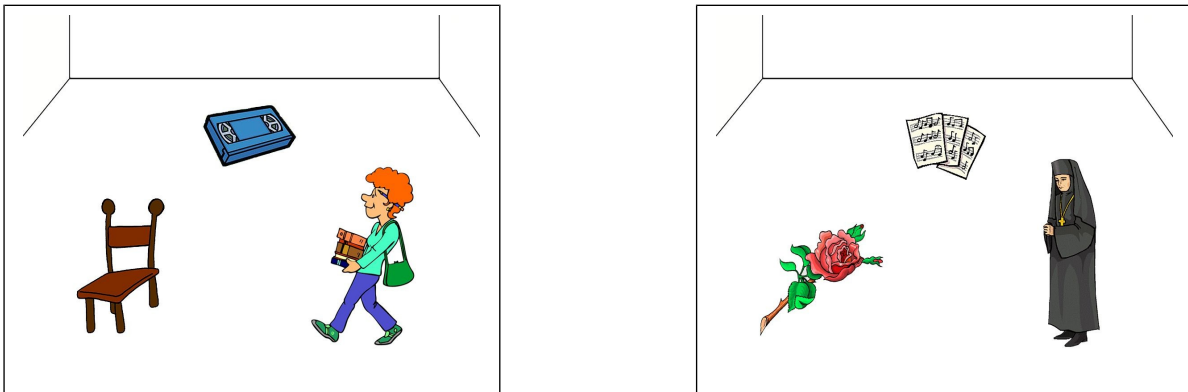

 Figure 5. Example stimuli for Experiment 2.

We are interested in participants' eye-movements at the time when the verb is heard. We predict that if participants initially adopted an incorrect main clause analysis on hearing *watched* in (3a), they would search for an entity that fits best as a direct object even though the combination of the subject noun and the verb's semantics indicates that the main clause analysis is highly implausible. Since none of the visual entities actually fit as a direct object of the verb in our material, they would end up with processing deadlock. On the other hand, when participants adopted a correct relative clause analysis on hearing *recorded* in (4a), they would immediately predict a by-agent phrase and search for a possible relative clause agent entity (*nun*). It has been shown that the by-agent is a part of the lexical representation of a verb even if it is not explicitly expressed (Mauner & Koenig, 1999), suggesting that such anticipatory eye-movements may

occur at the verb irrespective of the actual presence of the by-agent phrase. If we assume that people use lexical frequency information immediately at the verb to predict an upcoming syntactic structure, they should correctly analyse the relative clause and predict an upcoming relative clause agent immediately for high PastP verbs, but not for low PastP verbs. Therefore, we expect little or no difference in eye-movements between the reduced relative clause (ambiguous) condition and unreduced (unambiguous) condition for high PastP verbs (4a and 4b) but there should be a difference between the two conditions for low PastP verbs (3a and 3b).

Finding no difference between the two high PastP conditions at the verb would suggest that participants ruled out a main clause analysis at the earliest stage of processing and would therefore constitute evidence for the predictive usage of verb frequency information. If, however, the facilitative effect of high PastP verbs is due to later integration rather than early prediction, then participants should always construct a main clause structure and experience difficulty by not finding a thematically legitimate direct object, i.e., there should be no interaction between the type of relative clause (reduced vs. unreduced) and PastP bias (high vs. low).

Method

Participants

Twenty-nine new participants were recruited from the same population as in Experiment 1. One participant was excluded due to recording difficulty.

Materials

Twenty-four items were constructed, each of which consisted of a spoken sentence and a visual scene. For selecting verbs, we conducted two normative studies using WebExp experimental

3

software (Keller, Gunasekharan, Mayo, & Corley, 2009): a sentence completion study and a plausibility rating study. Thirty items using fifteen verbs for each condition (high PastP and low PastP) were created. Based on the results from these two studies, we selected the twelve best items for each group (i.e., 24 verbs in total). Fourteen of these verbs (7 verbs for each group) overlapped with the material of Trueswell (1996) and 10 verbs were new. Table 5 shows the mean verb duration, the mean onset of the preposition *by*, and that of the head noun in the *by*-phrase. As in Experiment 1, we checked whether different types of verbs differed in the length using an independent t-test. The results showed that the difference was not significant in the ambiguous condition ($t(21.15) = 1.50$, $p = 0.15$) but marginal in the unambiguous condition ($t(21.78) = 1.84$, $p = 0.08$).

Table 5. The mean duration of the relative clause verb and the mean onsets of the preposition *by* and of the head noun of the by-phrase, and the main verb in milliseconds. RC verb lengths in characters are given in brackets.

	Verb duration	Preposition by	Noun
Low reduced	791 (7.3)	1408	1766
High reduced	884 (8.4)	1500	1836
Low unreduced	739 (7.3)	1316	1697
High unreduced	843 (8.4)	1435	1778

Sentence completion study

Forty-four participants took part in the study. Each participant saw fifteen items with 28 fillers. The order of presentation was randomized for each participant. They were given sentence fragments that included a subject noun phrase and a verb and were asked to create a grammatical continuation of the fragment. The participants read instructions at the beginning of their session and went through a practice session with two example items.

The results showed that the sentence fragments with high PastP verbs were completed as past participle 84.6% of time, whereas those with low PastP verbs were 51.3% of time. The LME analysis on the logit of relative clause completions against all other completions showed that the difference is significant (coefficient = 2.71, z = 3.58, p < .001). Table 6 below shows the mean log ratio of past participle frequency and past tense frequency calculated from the results of our sentence completion study and from the written part of the British National Corpus (frequency data obtained from ftp://ftp.itri.bton.ac.uk/bnc/).

Table 6. The mean log ratio of past participle frequency and past tense frequency for the results from our completion study and BNC (standard errors in brackets).

	Low-PastP Verbs	High-PastP Verbs
Log(VVN/VVD) from Completion Study	-.001 (.42)	2.01 (.35)
Log(VVN/VVD) from BNC	-.70 (.18)	1.45 (.16)

Plausibility rating study

Thirty-six participants took part in the study. Each participant saw 15 items with 28 fillers. The order of presentation was randomized for each participant. Our interest is to see if the plausibility of possible structures given our experimental items differs between the two groups of verbs. We tested three conditions that correspond to three possible ways of assigning thematic roles to the three visual entities depicted in a scene. The first condition shows the correct syntactic analysis assigning agent and patient roles to the correct entities (5a, Correct role assignment). The second condition shows a main clause misanalysis that assigns an agent role to the sentence-initial noun and a patient to the entity that is an agent in the correct analysis (5b, Incorrect role assignment I).

The third condition shows a main clause misanalysis that assigns an agent role to the sentence-initial noun and a patient to the distractor (5c, Incorrect role assignment II).

5a. The student watched the video. (Correct role assignment)

5b. The video watched the student. (Incorrect role assignment I)

5c. The video watched the chair. (Incorrect role assignment II)

Participants saw a sentence from one of the three conditions and were asked to judge how plausible the sentence is. The rating ranged from 0 for 'very implausible' to 5 for 'very plausible'. The sentences were presented together with the picture entities that were later used in the experimental material in order to avoid any uncertainty or ambiguity about the referents. As in the sentence completion study, the participants read instructions at the beginning of their session and went through a practice session with two example items. Table 7 shows the mean z-scores for each condition for each verb group.

Table 7. The mean z-scores for each condition and each group of verbs (standard errors in brackets).

	Low-PastP	High-PastP	Total
Correct role assignment	1.19 (.14)	1.18 (.06)	1.19 (.08)
Incorrect role assignment I	-.69 (.07)	-.43 (.13)	-.56 (.08)
Incorrect role assignment II	-.62 (.06)	-.63 (.11)	-.63 (.06)

We conducted an LME analysis including the three sentence types (Sentence Type) and two verb groups (Verb Type; low PastP vs. high PastP) as predictor variables along with participants and items as random factors in the model. The results showed a main effect of Sentence Type, suggesting that the two incorrect sentence types are judged significantly worse in the rating than

the correct type (coefficient = -1.62, $t = 11.51$, $p < .001$ for Incorrect role assignment I and coefficient = -1.81, $t = 12.91$, $p < .001$ for Incorrect role assignment II). Importantly, however, there was no main effect of Verb Type (coefficient = 0.01, $t = .06$, *n.s.*) nor an interaction with either level of Sentence Type (coefficient = -0.27, $t = 1.35$, *n.s.* with Incorrect role assignment I and coefficient = 0.00, $t = .00$, *n.s.* with Incorrect role assignment II). In fact, taking out Verb Type and its interactions with Sentence Type improved the model fit, showing a more pronounced effect for Sentence Type (coefficient = -1.75, $t = 17.55$ for Incorrect role assignment I and coefficient = -1.82, $t = 18.19$ for Incorrect role assignment II). The results suggest that between the two groups of items, there was no difference in terms of plausibility of the possible interpretations with our visual stimuli.

Design and procedure

The experiment included 24 experimental items and four versions of each item were prepared for the four experimental conditions exemplified by (3, 4). Each participant saw only one of the four conditions of each item and the same number of items in each condition. The 48 filler items that were used in Experiment 1 were also included. As in Experiment 1, the materials were presented in a randomized order with the constraint that each experimental item was preceded by at least one filler item. At least two fillers appeared at the beginning of each experimental session. We again included six written prompts following fillers asking participants to describe the previous scene. The experimental procedure was otherwise identical to that in Experiment 1. Each experimental session took 30 to 40 minutes to complete.

Data analysis and results

Our main interest is the looks to the agent entity for the relative clause on hearing the verb, which is ambiguous between a main verb and relative clause verb. However, it is possible that participants would make initial gazes to the relative clause agent entity not because they predicted a relative clause but because they predicted a main clause and check its thematic fit as a direct object. This would make an analysis of the saccade probability (as in Experiment 1) somewhat obscure. Thus, we only present analyses on the logit of gazes in this experiment. As in Experiment 1, we used LME models for the analysis. We included Verb Type (low PastP vs. high PastP) and Clause Type (Reduced vs. Unreduced) as main predictor variables, Time Window as a covariate and participants and items as random effects (items were treated as a between-item variable). Below, in cases where an interaction of fixed effects (but not for any interactions with a covariate) is significant, we report pairwise comparisons. The contrasts are computed using 95% confidence intervals derived from the relevant contrast in the LME model. Our analysis treats the reduced/high PastP condition as a baseline against which the other conditions and the interaction are compared. Note that interactions are interpreted in the same way as in standard Anova models: an interaction of Verb Type and Clause Type indicates that the unreduced/low PastP condition is significantly different from what is expected by the additive contribution of the two predictors, compared to the baseline (reduced/high PastP). We optimized the model-fit in the same way as Experiment 1. Like Experiment 1, we report coefficients, t-values, 95% confidence intervals, and significance levels for all fixed factors and the interactions that remained in the optimized model.

Gaze logit analysis

The region of our primary interest is the relative clause agent entity. To examine the influence of the verb, like in Experiment 1, we excluded the trials on which participants were already looking

at the the relative clause agent entity at the onset of the verb. Though the number of excluded trials was quite large (27% of all the trials), it did not differ across conditions ($\chi^2 = 0.03$, $df = 1$, $p = 0.9$). Figure 6 plots the gaze probability to the relative clause agent entity aggregated for each verb type and clause type in each 20 ms time bin over 1500 ms, starting from the verb onset.
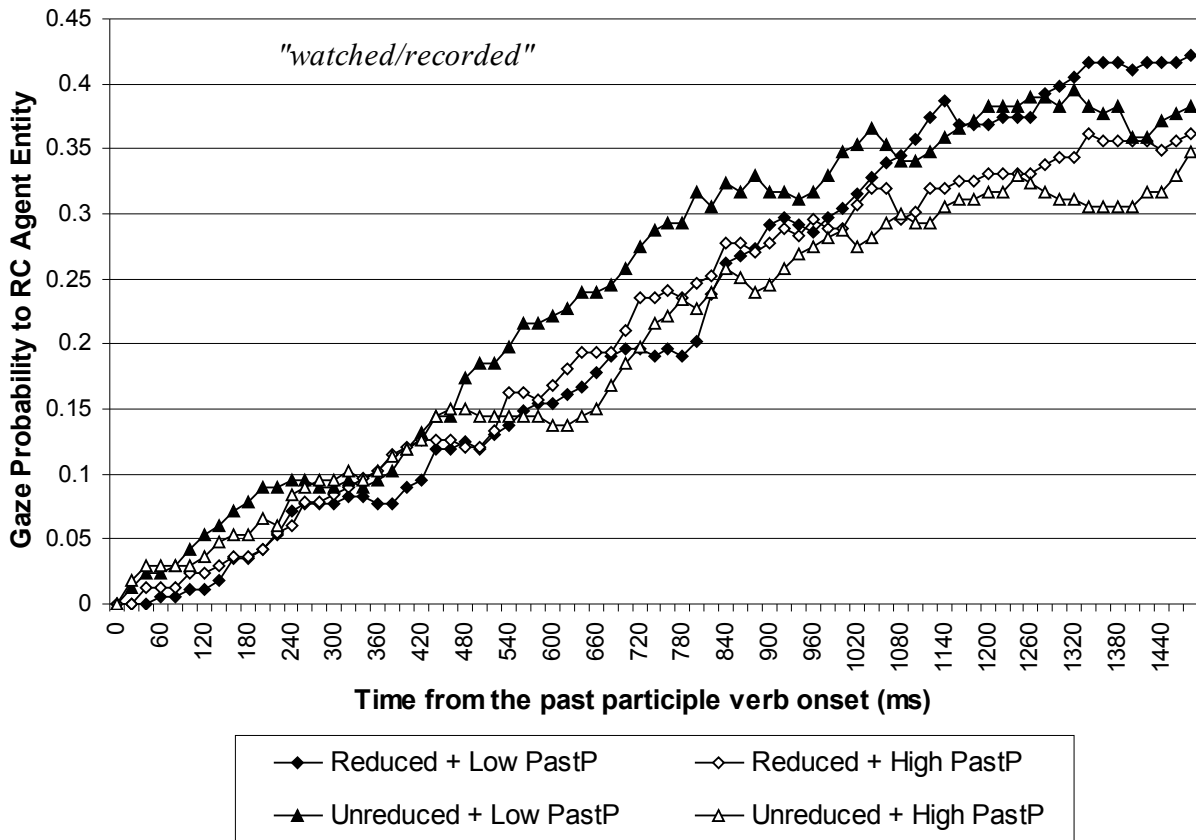

Figure 6. Gaze probability to the relative clause agent entity aggregated for each verb type and clause type condition, starting from the verb onset. The two vertical lines mark the mean verb offset and the onset of the postverbal preposition *by* respectively.

We first calculated the logit of gaze probabilities for each 100 ms time bin using the same function as in Experiment 1. Again, to avoid possible influence of the pause following the verb in the speech, we restricted the time interval for the analysis to 100 – 800 ms following the verb onset (mean duration of the verb was 814 ms, SD = 150). Note that since Verb Type is between-

items manipulation, any difference in looks to the relative clause agent entity between the two conditions could be due to the difference in visual properties of pictures used for each verb group and bears little theoretical interest.

Table 8. Analysis of gaze logits to the relative clause agent entity for 700 ms interval from 100 ms to 800 ms following the verb onset.

Factor	coefficient	t	CI	sig.
(Intercept)	-3.01			
Clause Type	0.02	1.91	0.02	
Verb Type	-0.01	0.43	0.06	
Time Window	0.15	12.25	0.02	<.001
Clause Type × Verb Type	-0.04	3.00	0.02	<.01
Clause Type × Time Window	0.00	0.26	0.02	
Verb Type × Time Window	-0.01	0.91	0.02	
Clause Type × Verb Type × Time Window	-0.03	2.17	0.02	<.05

*p<0.05, **p<0.01, ***p<0.001

Table 8 summarizes the results from the model. We observed a main effect of Time Window. Participants looked increasingly more at the relative clause agent entity over the time period. Most importantly, there was a significant interaction between Clause Type and Verb Type. We now explore the pattern of the interaction by pairwise comparisons. The difference between the empirical means for the Clause Type contrast for low PastP verbs was 0.12 whereas the difference for the Clause Type contrast for high PastP verbs was 0.03. Given the confidence interval for the interaction (0.02), both differences are reliable. However, the small difference for high PastP verbs is actually in the opposite direction from what was expected: higher logit for the reduced condition than the unreduced condition. This is most likely due to more looks to the distractor entity in the unreduced condition as will be seen later. Thus, the predicted difference caused by structural ambiguity was observed only for the low PastP verbs.

3

Furthermore, we also observed a three-way interaction between Clause Type, Verb Type, and Time Window. As Figure 6 shows, this suggests that the interaction between Clause Type and Verb Type became greater over the course of the time interval.[iii]

Like in Experiment 1, we examined the looks to the subject entity and to the distractor entity (as shown in Figure 7), in order to ensure that the effect we observed is unique to the target entity, i.e., it is not due to an artefact that concerns all the entities in the display. We analyzed looks to the subject and distractor entities in the same way as the looks to the relative clause agent entity.
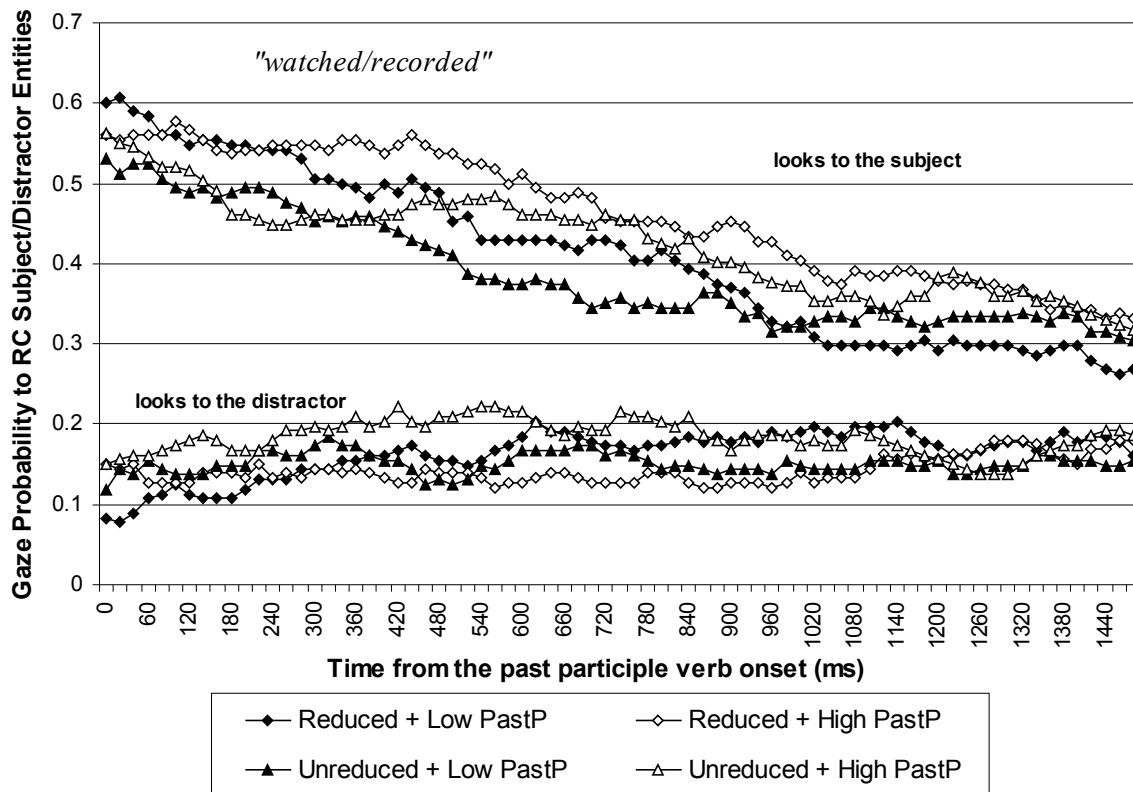

Figure 7. Gaze probability to the subject and distractor entities aggregated for verb type and clause type condition, starting from the verb onset.

Table 9. Analysis of gaze logits to the subject entity for 700 ms interval from 100 ms to 800 ms following the verb onset.

4

Factor	coefficient	t	CI	sig.
(Intercept)	2.46			
Clause Type	-0.06	4.34	0.03	<.001
Time Window	-0.18	12.25	0.03	<.001

*p<0.05,

The LME analysis on the gaze logits to the subject entity (see Table 9) shows main effects of Clause Type and Time Window. There are less looks to the subject entity in the unreduced relative clause condition than in the reduced relative clause condition and the looks to the subject entity decreased over the time period. Importantly, there was neither an effect of Verb Type nor its interaction with any other variables, which demonstrates that verb frequency information did not affect the looks to the subject entity. The difference between the unreduced and reduced relative clause condition is most likely due to the extra time that existed before hearing the past participle verb for the unreduced condition (i.e., the time for *that was*). In fact, the onset of the past participle verbs was on average 476 ms later for the unreduced condition compared to the reduced condition.

Table 10. Analysis of gaze logits to the distractor entity for 700 ms interval from 100 ms to 800 ms following the verb onset.

Factor	coefficient	t	CI	sig.
(Intercept)	-2.94			
Clause Type	0.03	2.62	0.03	<.05
Verb Type	0.02	0.59	0.06	
Clause Type x Verb Type	0.04	3.10	0.03	<.01

*p<0.05, **p<0.01, ***p<0.001

The results of the LME analysis on the gazes to the distractor entity are given in Table 10. They show a main effect of Clause Type. There are more looks to the distractor entity in the unreduced than reduced relative clause condition. More importantly, the interaction between Clause Type and Verb Type was significant. The empirical mean difference between the reduced and unreduced condition for the low PastP verbs was 0.01 whereas it was 0.15 for the high PastP verbs. Given the confidence interval for the interaction (0.03), only the latter difference was reliable. This is therefore the inverse of what we observed for the looks to the relative clause agent entity, where a significant difference between the reduced and the unreduced condition was found for the low PastP, but not for the high PastP verbs. On hearing the high PastP verb, participants looked less at the subject entity in the unreduced condition than in the reduced condition but looked at the relative clause agent entity to a similar extent in both conditions. The difference for the looks to the subject entity was reflected in the looks to the distractor entity; they looked more at the distractor entity in the unreduced condition than in the reduced condition. We did not, on the contrary, observe such a difference in looks to the distractor entity with the low PastP verb. This is possibly because the unreduced/high PastP condition is the least difficult case as the structural cue and the verb frequency information both support a relative clause reading and participants may have considered a possibility for some other continuations with the distractor entity (e.g., *The song that was recorded about the rose…*). In any case, the looks to the distractor entity do not account for the difference in the looks to the relative clause agent entity between the reduced and unreduced conditions for low PastP verbs and for no difference for high PastP verbs.

We also analyzed the gazes following the preposition *by* to see whether the past participle frequency information would have any effect after hearing the disambiguating information, i.e., the preposition *by*. Figure 8 shows the logit of gazes to the relative clause agent entity aggregated

4

for each verb type and clause type in each 20 ms time bin over 1500 ms from the preposition onset. We did not exclude the trials in which participants were looking at the relative clause agent entity at the onset of the preposition because we are interested in how the eye-movements, including anticipatory or responsive, reflected the by-agent information.
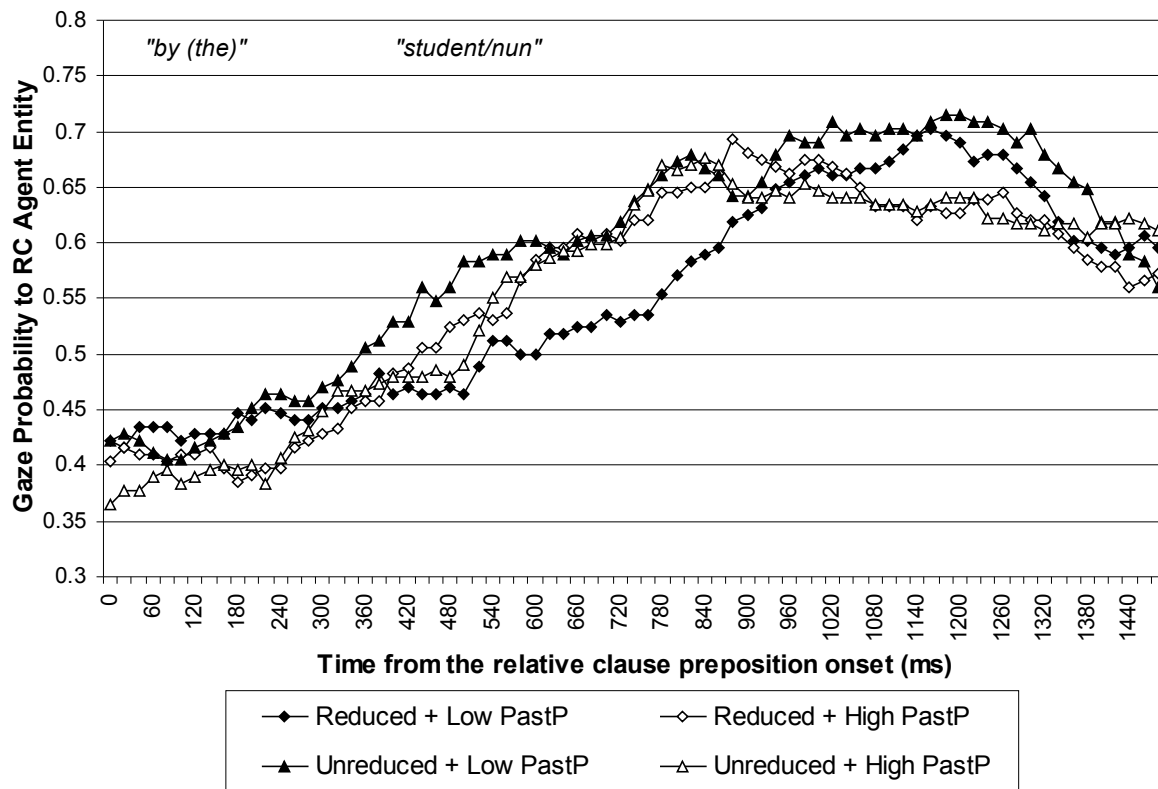

Figure 8. Gaze probability to the relative clause agent entity for each verb type and clause type condition, starting from the onset of the preposition *by*. The average onset of the noun following the preposition is marked by a vertical line.

Note that during this time period, participants heard not only the preposition but also the agent complement noun (mean complement noun onset is 355 ms, SD = 60 ms). However, the noun was lexically identical for the reduced and unreduced conditions and thus any difference cannot be due to the processing of this complement noun. We selected the 800 ms time interval from 100

ms to 900 ms following the preposition onset. As seen in the graph, the gazes to the relative clause agent entity peak around 900 ms following the preposition onset and gradually decrease after that point. Thus, inclusion of any time window after 900 ms would jeopardize the linearity assumption of LME. We entered Verb Type and Clause Type as predictor variables, and Time Window as a covariate. Table 11 summarizes the results from the optimized model.

Table 11. Analysis of gaze logits for 800 ms interval from 100 ms to 900 ms following the preposition onset.

Factor	coefficient	t	CI	sig.
(Intercept)	-2.11			
Clause Type	0.04	2.12	0.03	<.05
Verb Type	-0.01	0.25	0.08	
Time Window	0.18	10.30	0.03	<.001
Clause Type x Verb Type	-0.04	2.31	0.03	<.05
Verb Type x Time Window	0.04	2.26	0.03	<.05

*p<0.05, **p<0.01, ***p<0.001

We observed main effects of Clause Type and Time Window. While there was no effect of Verb Type, the interaction between Verb Type and Time Window was significant. The coefficient suggests that the looks to the relative clause agent entity increased more greatly over time with high PastP verbs than with low PastP verbs. However, importantly, we found a significant interaction between Verb Type and Clause Type. We again explored the pattern of the interaction with pairwise comparisons. The difference between the empirical means for the Clause Type contrast for the low PastP verbs was 0.15 whereas the difference for the Clause Type contrast for the high PastP verbs was 0.01. Given the confidence interval for the interaction (0.03), only the former difference is reliable. Again, we found a difference between the reduced and unreduced conditions for the low PastP verbs but not for high PastP verbs, confirming that participants showed no evidence for reanalysis on hearing the disambiguating material with high PastP verbs.

Discussion

The results clearly showed that when a low PastP verb was heard, participants made more looks to the relative clause agent entity in the unreduced condition than in the reduced condition. In the reduced condition, participants presumably followed an incorrect interpretation due to structural ambiguity (garden-path), initially adopting a main clause analysis for the low PastP verbs, and thus experienced difficulty in finding a thematically legitimate direct object. On the contrary, when a high PastP verb was heard, there was no difference in looks to the relative clause agent entity between the reduced and unreduced relative clause conditions. In other words, there was no evidence for any cost due to structural ambiguity from the earliest stage of processing the verb. The results therefore demonstrate that participants immediately adopted the correct relative clause analysis if verb-specific frequency information and plausibility of the subject noun supported it, and thus looked at the relative clause agent entity that they predicted based on this frequency information.

Furthermore, we found a similar ambiguity cost with low PastP verbs when the disambiguating prepositional phrase was encountered. This suggested that although visual information clearly indicates that the main-clause interpretation is not possible (as confirmed by our normative test), participants had not fully resolved the structural ambiguity with the low PastP verbs until they encountered the disambiguating information. Again, we found no such cost for high PastP verbs following the preposition onset; participants immediately adopted a correct relative clause analysis with these verbs and maintained the analysis without considering the alternative main-clause analysis throughout processing of the relative clause.

Taken together, the results demonstrated that participants accessed lexically-specific frequency information immediately on encountering the verb and used it to predict the correct

relative clause structure before any postverbal information was encountered. The results showed no evidence for syntactic misanalysis (or garden-path effect) with high PastP verbs, which contrasted with low PastP verbs with which clear processing difficulty associated with structural ambiguity was observed. The difference observed between low PastP verbs and high PastP verbs therefore is due to prediction of the correct syntactic structure based on lexically specific frequency information and it cannot be explained in the terms of integration of disambiguating material, given that the effect occurred early (at the verb), and that there was no evidence for syntactic misanalysis for high PastP verbs throughout the sentence.

General Discussion

Experiment 1 showed that participants used verb-specific subcategorization information to make predictions about whether the verb is followed by a direct object or not. Also, similar anticipatory eye-movements were observed on hearing a preposition following the verb, demonstrating that the participants predicted an upcoming complement of the to-be-built prepositional phrase. Experiment 2 demonstrated that participants used information about verb form frequencies (how often a verb occurs in the past participle form) for predictions about the correct structure. The temporally ambiguous verb in a reduced relative clause caused a garden-path effect only for low PastP verbs, but not for high PastP verbs. The results from our two experiments together demonstrated that comprehenders can extract lexically-specific syntactic information incrementally and use it to predict upcoming linguistic material. These effects were observed at the earliest stage of processing the target word and thus indicate that verb-specific information is used immediately to anticipate upcoming syntactic structure. This in turn can be used to predict lexical elements compatible with the anticipated structure, which manifests itself as anticipatory eye-movements in our visual word setting.

4

The current study contributes to our understanding of an important aspect of human sentence processing, viz., prediction. There is a considerable body of evidence which suggests that people's predictions about upcoming linguistic material are driven by semantic information (e.g., Altmann & Kamide, 1999, Chambers et al., 2002) and discourse constraints (e.g., Kaiser and Trueswell, 2004). Studies investigating syntactic prediction have focused on case marking and verb tense as triggers for prediction (Kamide et al., 2003a, b, Altmann & Kamide, 2007). These syntactic triggers are not lexically specific: case and tense marking is independent of the lexical items used. The present study, in contrast, focused on the use of information in prediction that is specific to individual lexical items: the subcategorization frame of the verb, and the frequency with which the verb occurs as a past participle. In two experiments, we demonstrated that such lexically specific information is used immediately and can influence the syntactic analysis at the earliest point in processing (i.e., when the verb is encountered). Therefore, our results argue in favor of a prediction-based account and against an account in which verb-specific information merely facilitates the integration of disambiguating material further downstream.

Our findings raise the question of why many previous studies have failed to observe an immediate effect of verb-specific information (e.g., Ferreira & Henderson, 1990; Kennison, 2001; Mitchell, 1987; Pickering et al., 2000; Pickering et al., 2003). These studies typically found processing disruption at the postverbal NP, which is implausible as the verb's direct object, even when the verb prefers an alternative structure that requires no direct object, such as a sentence complement (e.g., *The young athlete realized her exercises...*, Pickering et al., 2000). Previous studies interpreted these results as evidence against the influence of lexical frequency information on the initial parse. However, such an interpretation is debatable because the effects in these studies were observed when readers encountered the postverbal material and tried to integrate it into the representation they had already built, rather when they encountered the verb itself as in

4

our study. As they investigate the processing of the verb *plus* the postverbal material, previous studies are only able to provide indirect evidence regarding the influence of lexically-specific verb information on the initial syntactic analysis. What these studies instead suggest is that when an NP is encountered directly following a verb, lexically specific information tends to be overridden by the strong tendency to analyse it as a direct object. We suppose that this is probably because most verbs, including intransitive verbs, can optionally take a direct object (with the exception of unaccusative verbs such as *arrive*, see Staub & Clifton, 2006) and the main clause NP analysis occurs most frequently among all possible analyses.

In addition to this, there is evidence that the structural biases of a verb can differ with respect to the sense of the verb. As an extreme example, take again *realize:* in the 'know, cognize' sense of the verb, only the sentential complement is possible, while in the 'make, create' sense, only the NP complement is possible. Similar sense-specific biases can be observed across a whole range of structually ambiguous verbs, and failure to take sense distinctions into account can explain why some studies in the literature have failed to find evidence for the immediate use of structural preferences (Hare, McRae, & Elman, 2004).

In fact, Trueswell (1996) demonstrated that when the sentence initial subject NP is a good agent, which is strongly associated with a main clause analysis, lexical frequency information itself cannot reverse the preference for a relative clause structure (his Experiment 2). This is good evidence that readers take not only lexically-specific information but also other relevant constraints (such as plausibility of the subject) into account when selecting a syntactic analysis. What our study clearly demonstrated is that comprehenders use lexically specific information as one reliable source of information among others to make predictions about an upcoming syntactic structure, and that this information is used as soon as possible, i.e., at the lexical items itself.

Furthermore, Trueswell's (1996) study also indicates that prediction is at work throughout the processing of all the linguistic input, not only at the verb. That is, when participants heard a first noun phrase such as *the nun* in Experiment 1, they are likely to interpret it as an agent and expect a verb that permits a main clause analysis. The following input in our experimental materials, both for transitive and intransitive verbs, did not contradict the prediction. However, in Experiment 2, when participants heard a first noun phrase that was good theme and bad agent such as *the song*, they are unlikely to expect a verb that normally requires an animate agent such as an action verb (e.g., *kick*) or state verb (e.g., *believe*). Instead, it is more likely that they would expect verbs that can take an inanimate noun as the subject such as copular verbs (e.g., *sound*) or passive verbs (e.g., *was played*). Such expectations mean that low PastP verbs are incompatible with predicted input, while high PastP verbs as compatible with it – the former are strongly associated with the unexpected main clause, whereas the latter are associated with the expected passive. The prediction of a verb at the sentence initial NP was not investigated by the current study and it is still an empirical question whether and how readers selectively restrict the set of upcoming verbs. However, Experiment 2 failed to find any evidence for the more frequent but incorrect main clause analysis for high PastP verbs, even at the earliest point of the processing. This strongly suggests that predictive processes are likely to be operating on a word-by-word basis. This view is consistent with processing models that estimate syntactic predictions on a word-by-word basis (Hale, 2001, Levy, 2008).

An important issue remains: Should the lexically specific subcategorization information that drives the anticipatory eye-movements in Experiment 1 be seen as entirely syntactic or partly semantic? We regard it as primarily syntactic as far as the nature of the difference between the two groups of verbs is concerned (i.e., the distinction was based on the entries of subcategorization frames). However, this issue is complicated by the fact that as some linguistic

theories assume that subcategorization frames are not entirely separate from semantic information. Thematic role semantics is often assumed to be lexical as well, and at least partially related to subcategorization information (Kaplan & Bresnan, 1982; Chomsky, 1981). According to this view, the reason why a verb like *punish* takes two argument roles is partly due to its semantics; an act of punishing includes somebody to punish (i.e., agent) and somebody to be punished (patient). Similarly, an act of disagreeing can include both somebody to disagree and somebody or something to be disagreed with, although the latter information does not need to be explicitly expressed. It suggests that the semantic portion of lexically specific subcategorization information is not sufficient to cause the differential eye-movements in Experiment 1. Therefore, what causes the difference is a syntactic rule that specifies the verb *punish* takes a direct object but *disagree* does not. This is contrary to previous studies such as Altmann and Kamide (1999) in which there is no syntactic motivation to distinguish between the verb *move* and *eat* as both verbs license a direct object (something to be moved or eaten).

Another important issue is the role that visual information plays in the pattern of results observed in the current study. It is reasonable to assume that visual information, provided together with linguistic information, could have facilitated the predictive processes observed in the current study, although it is hard to determine exactly how much of an influence it had. We can think of two contrasting views regarding the role of visual information in visual world experiments investigating prediction. One view is that predictions of upcoming information are primarily driven by linguistic input and people search for entities in the visual scene that are compatible with the predictions. The other view is that the information supplied by the visual entities is already processed before predictions are made and listeners check the consistency or fit between the visual entities and the most appropriate sentence continuation. Although the current data cannot distinguish between the two views, previous research has provided relevant evidence.

The results from reading studies, in the absence of a visual scene, demonstrate that prediction can be driven purely on the basis of linguistic information (see Introduction for details). On the other hand, a series of experiments conducted by Altmann and his colleagues indicate that listeners sometimes make over-specified prediction because of the extra visual information. For example, Altmann and Kamide (2010) report that participants predict an empty glass after hearing *the boy has drunk,* but it is not clear whether such detailed predictions would occur if there is no contrast present in the visual scene between an empty glass and a full glass**.** Taken together, both types of studies suggest that prediction is at work even in the absence of visual information, but can be encouraged by its presence.

We will now turn to the implications of our results for models of sentence processing. As we argued in the introduction, surprisal-based models are the only current models that can account for evidence that human language processing is predictive. Such models are based on Hale's (2001) hypothesis that processing difficulty is correlated with the change in the probability distribution over possible analyses from one word to the next: words that trigger a small change in the probability distribution (i.e., that are predictable) are easy to process, while words that trigger a large change in the distribution cause processing difficulty.

The surprisal approach does not, however, provide an insight into the type of information that drives prediction. In fact, as Levy (2008) shows, surprisal only requires a language model (in the speech processing sense of the term), i.e., a probabilistic model that assigns probabilities to word sequences. This can be achieved by a range of probabilistic devices, including simple recurrent networks or *n*-gram models (which only use word transition probabilities), hidden Markov models (which use part-of-speech transition probabilities and word emission probabilities), and structured language models that employ probabilistic grammars to compute probabilities for word sequences.

Most proponents of surprisal assume that surprisal values are computed based on a probabilistic grammar (Hale, 2001, Levy, 2008, Demberg & Keller, 2008, 2009, Boston, Hale, Kliegl, Patil & Vasishth, 2008). This means that prediction is driven by syntactic information, i.e., by the difference in the probability distribution over possible syntactic structures at the previous word compared to the current word. However, Frank (2009) advocates a view of surprisal in which prediction happens strictly at the word level. In this approach, surprisal values are estimated using a simple recurrent network (SRN, Elman, 1991) as the difference in the probability distribution over possible word sequences at the previous word compared to the current word. This word-based notion of surprisal does not have access to syntactic information (other than what is encoded in word sequences).

In this context, it is important to bear in mind that all existing surprisal-based models are designed to account for language processing in the abssence of a visual context. Few attempts have been made to model the interaction of linguistic and visual processing in a single model, which would then be able to account for visual world data such as the one resported in the present article. A notable exception is the Coordinated Interplay Account (CIA) proposed by Knoeferle and Crocker (2007) and Mayberry, Knoeferle, and Crocker (2009). The CIA model is implemented as an SRN, which in addition to textual input also takes a representation of the events in a visual scene as input. Both types of input share a hidden layer, which enables the network to use the visual scene as a context for the interpretation of the text, outputting an event-based interpretation of the utterace. The CIA can be seen as extension of a suprisal model, in which suprisal is computed incrementally by the SRN (as in Frank's, 2009, work), but is modulated by a representation of the visual context. (For conceptually similar ideas, which are however not computationally implemented, see Altmann & Mirkovic, 2009, and Huettig, Olivers, & Hartsuiker, 2011).

The results presented in the current paper support a surprisal view of prediction on a general level; Experiment 2 shows that frequency information (verb form frequencies) are used for generating predictions, which is what we expect under a probabilistic approach to prediction, which is shared by all surprisal models, syntax-based and word-based ones. The results of Experiment 1, however, can be used to discriminate between the two model types: most syntax-based models of surprisal incorporate subcategorization information. An example is Demberg and Keller's (2009) model, which computes surprisal estimates using a variant of tree-adjoining grammar (TAG). In this approach, the TAG lexicon entry of a verb generates predictions in the form of open substitution nodes with which upcoming arguments of a verb have to be connected. The differential behavior of transitive and intransitive verbs that we saw in Experiment 1 is therefore expected, as these two verb types have lexicon entries in TAG which differ in terms of their substitution nodes. Only transitive verbs have a substitution node for a postverbal argument, and thus only these verbs trigger aniticipatory looks to an entity of such an argument in a visual scene.

However, a purely word-based view of surprisal such as Frank's (2009) is not able to account for the findings of Experiment 1. Subcategorization information is not encoded by his SRN, which is only able to model word sequences. The fact that transitive verbs such as *punish* trigger anticipatory looks, while intransitive verbs such as *disagree* do not is not expected under this account. In both cases, the SRN would predict the next word, which presumably is the start of an NP in the case of *punish* (e.g., *the*), while in the case of *disagree* it is probably the start of a prepositional phrase (e.g., *with*). This type of prediction does not explain why we observe anticipatory looks to a potential argument of the verb in the case of *punish*, but not in the case of *disagree*. Such a prediction is only possible if we have a notion of verb-argument relationship as it is afforded by TAG or in general by a grammatical framework that is able to represent verb

sucategorization frames. Note that it is conceivable, however, that an SRN-based model could be extended to include sucategorization information, perhaps along the lines of Mayberry et al.'s (2009) CIA model, described above. In this case, the event-based representation used by the CIA could be augmented with subcatetorization information.

Conclusions

The aim of this paper was to contribute to our understanding of prediction in sentence processing, which have generated a lot of interest in the recent sentence processing literature. In particular, we asked whether prediction makes use of lexically specific syntactic information, such as subcategorization information and verb form frequencies. We presented the results of two visual world experiments. Experiment 1 examined whether verb subcategorization information is used for predicting a direct object, by comparing transitive verbs (e.g., *punish*) to intransitive verbs (e.g., *disagree*). The results showed that comprehenders use subcategorization information at the verb to launch predictive eye-movements to a visual entity of plausible direct object in the case of transitive verbs, but not in the case of intransitive verbs. Experiment 2 examined whether verb frequency information is used for predicting a relative clause by contrasting verbs that are infrequent in the past participle form (e.g., *watch*) with ones that are frequent in that form (e.g., *record*). The results demonstrated that participants immediately adopted a correct relative clause analysis with verbs of high past participle frequency and maintained it without considering the incorrect main clause analysis. For verbs of low past participle frequency, however, a garden-path effect was observed. Taken together, these two experiments demonstrate that comprehenders access lexically specific syntactic information for prediction and that this information can be used immediately in the parsing process, rather than only contributing to integration of disambiguating information into the structure already built independently of that information. We suggest that

our findings provide support for surprisal-based models of sentence processing, in which processing difficulty is related to how well the current input predicts upcoming linguistic material. Specifically, our results favor a syntax-based version of surprisal, as the use of subcategorization information in prediction is hard to reconcile with a purely word-based formulation of surprisal.

References

Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Altmann, G. T. M. and Kamide, Y (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502–518

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediciton in human sentence processing. *Cognitive Science*, *33*, 583-609.

Andersson, R. (2009). The Visual World paradigm in settings of high information load. Proceedings of the Second Conference of the Swedish Association for Language and Cognition, Stockholm University, Sweden.

Apel, J., Knoeferle, P., & Crocker, M.W. (2007). Processing parallel structure: Evidence from eye-tacking and a computational model. *Proceedings of the 2nd European Cognitive Science Conference* (pp. 125-131). Delphi, Greece.

Arai, M., van Gompel, R. P. G. & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*, 218-250.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457-474.

Bates, D. (2010). Linear mixed model implementation in lme4. (R package version 0.999375-33).

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity - evidence for cue trading relations. *Journal of Memory and Language*, *30*, 644-663.

Boland, J. (2005). Visual Arguments. *Cognition*, *95*, 237–274.

Boston, M. F., Hale, J. T, Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2*, 1-12.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H. & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*, 30-49.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Clifton, C., Frazier, L. & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *23*, 696-708.

DeLong, K. A., Urbach, T. P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121.

Demberg, V. & Keller, F. (2008). Data from Eye-tracking Corpora as Evidence for Theories of Syntactic Processing Complexity. *Cognition, 109*, 193-210.

Demberg, V. & Keller, F. (2009). A Computational Model of Prediction in Human Parsing: Unifying Locality and Surprisal Effects. *Proceedings of the Annual Conference of the Cognitive Science Society*, Amsterdam, Netherlands.

Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning, 7*, 195-225.

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

Ferreira, F. & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 555-568.

Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In: N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1139-1144). Austin, TX: Cognitive Science Society.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L. (1991). *Exploring the architecture of the language-processing system.* In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 409-433). Cambridge, MA: MIT Press.

Frazier, L., Munn, A. & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*, 343-370.

Garnsey, S. M., Pearlmutter, N. J., Myers, E. & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58-93.

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

Grishman, R., Macleod, C. & Meyers, A. (1994). Comlex syntax: Building a computational lexicon. *Proceedings of the 15th International Conference on Computational Linguistics* (pp. 268-272). Kyoto, Japan.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159-166). Pittsburgh, USA.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*, 609-642.

Hare, M., McRae, K. & Elman, J. L. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes, 19*, 181–224.

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*, 138-150.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137-194.

Kaiser, E. & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147

Kamide, Y., Altmann, G. T. M. & Haywood, S. L. (2003b). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-156.

Kamide, Y., Scheepers, C. & Altmann, G. T. M. (2003a). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37-55.

Kaplan, R. M., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173-281). Cambridge, MA: MIT Press.

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of web experiments: A case study using the WebExp software package. *Behavior Research Methods, 41*, 1-12.

Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review*, *8*, 132-138.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*, 95-127.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye-movements. *Journal of Memory and Language*, *57*, 519–543.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627-645.

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

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

Matin, E., Shao, K. C. & Boff, K. R. (1993). Saccadic overhead - information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372-380.

Mauner, G. & Koenig, J.-P. (1999). Lexical Encoding of Event Participant Information. *Brain and Language*, *68*, 178-184.

Mayberry, M., Crocker, M., & Knoeferle, P. (2009). Learning to Attend: a connectionist model of situated language comprehension. *Cognitive Science*, *33*, 449-496.

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

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Language* (pp. 601-618). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pickering, M. J. & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*, 469-503.

Pickering, M. J., Traxler, M. J. & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447-475.

Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ, US, Prentice-Hall, Inc.

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

Staub, A. & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either ... or. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*, 425-436.

Steedman, M. (1996) *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.

Stowe, L. A., Tanenhaus, M. K. & Carlson, G. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, *34*, 319-340.

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language, 62*, 333-351.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566-585.

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

Tyler, L. K., & Warren, P. (1987). Local and Global Structure in Spoken Language Comprehension. *Journal of Memory and Language*, *26*, 638-657.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*, 443-467.

Wicha, N. Y. Y., Moreno, E. M. & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272-1288.

Appendix 1

Experimental items used in Experiment 1.

1a. Suddenly, the rugby player scolded the shop assistant.

1b. Suddenly, the rugby player shouted at the shop assistant.

1c. Suddenly, the rugby player shouted and the shop assistant scratched the piano.

2a. Suddenly, the cheerleader mocked the runner.

2b. Suddenly, the cheerleader screamed at the runner.

2c. Suddenly, the cheerleader screamed and the runner grabbed the cheque.

3a. Apparently, the tennis player attracted the model.

3b. Apparently, the tennis player smiled at the model.

3c. Apparently, the tennis player smiled and the model took the key.

4a. Clearly, the waiter disgusted the tourist.

4b. Clearly, the waiter coughed at the tourist.

4c. Clearly, the waiter coughed and the tourist ate ice cream.

5a. Obviously, the knight rescued the peasant.

5b. Obviously, the knight waved at the peasant.

5c. Obviously, the knight waved and the peasant hid the loaf.

6a. Politely, the bellboy greeted the actress.

6b. Politely, the bellboy bowed to the actress.

6c. Politely, the bellboy bowed and the actress received the balloon.

7a. All of a sudden, the inmate offended the judge.

7b. All of a sudden, the inmate frowned at the judge.

7c. All of a sudden, the inmate frowned and the judge threw the gloves.

8a. Abruptly, the carpenter embarrassed the ballerina.

8b. Abruptly, the carpenter giggled at the ballerina.

8c. Abruptly, the carpenter giggled and the ballerina hurled the book.

9a. Obviously, the reporter insulted the jockey.

9b. Obviously, the reporter laughed at the jockey.

9c. Obviously, the reporter laughed and the jockey kicked the bicycle.

10a. Apparently, the bandit slapped the hiker.

10b. Apparently, the bandit spat at the hiker.

10c. Apparently, the bandit spat and the hiker ate the cake.

11a. Publicly, the drama critic condemned the professor.

11b. Publicly, the drama critic complained to the professor.

11c. Publicly, the drama critic complained and the professor threw the hat.

12a. Obviously, the graduate student accused the opera singer.

12b. Obviously, the graduate student chatted to the opera singer.

12c. Obviously, the graduate student chatted and the opera singer smashed the mirror.

13a. Secretly, the soldier contacted the doctor.

13b. Secretly, the soldier looked at the doctor.

13c. Secretly, the soldier looked and the doctor was playing chess.

14a. Secretly, the gardener seduced the gymnast.

14b. Secretly, the gardener winked at the gymnast.

14c. Secretly, the gardener winked and the gymnast smoked a cigarette.

15a. Abruptly, the butcher blamed the vendor.

15b. Abruptly, the butcher growled at the vendor.

15c. Abruptly, the butcher growled and the vendor ate cheese.

16a. Apparently, the witch upset the prince.

16b. Apparently, the witch glared at the prince.

16c. Apparently, the witch glared and the prince threw the shoes.

17a. Apparently, the miner respected the mechanic.

17b. Apparently, the miner listened to the mechanic.

17c. Apparently, the miner listened and the mechanic was talking to the chair.

18a. All of a sudden, the wizard abducted the boy.

18b. All of a sudden, the wizard stared at the boy.

18c. All of a sudden, the wizard stared and the boy hid the hamburger.

19a. Unexpectedly, the wrestler distracted the nurse.

19b. Unexpectedly, the wrestler jumped at the nurse.

19c. Unexpectedly, the wrestler jumped and the nurse dropped the frying pan.

20a. Surprisingly, the nun punished the artist.

20b. Surprisingly, the nun disagreed with the artist.

20c. Surprisingly, the nun disagreed and the artist threw the kettle.

21a. Apparently, the pirate irritated the princess.

21b. Apparently, the pirate whispered to the princess.

21c. Apparently, the pirate whispered and the princess opened the umbrella.

22a. Unexpectedly, the director praised the swimmer.

22b. Unexpectedly, the director yelled at the swimmer.

22c. Unexpectedly, the director yelled and the swimmer wore the belt.

23a. Surprisingly, the queen rewarded the butler.

23b. Surprisingly, the queen cared for the butler.

23c. Surprisingly, the queen cared and the butler cleaned the picture.

24a. Clearly, the poet disappointed the librarian.

24b. Clearly, the poet moaned at the librarian.

24c. Clearly, the poet moaned and the librarian read the map.

Appendix 2

Experimental items used in Experiment 2.

Low PastP verbs

1a. The pheasant hunted by the prince was eating the apple.

1b. The pheasant that was hunted by the prince was eating the apple.

2a. The infant entertained by the clown was interested in the cake.

2b. The infant that was entertained by the clown was interested in the cake.

3a. The money wanted by the model was for buying the ring.

3b. The money that was wanted by the model was for buying the ring.

4a. The sofa scratched by the boy was replaced with the office chair.

4b. The sofa that was scratched by the boy was replaced with the office chair.

5a. The videotape watched by the student was found under the chair.

5b. The videotape that was watched by the student was found under the chair.

6a. The house visited by the rock star was too small for the piano.

6b. The house that was visited by the rock star was too small for the piano.

7a. The wheelchair pushed by the bellboy was heavier than the chainsaw.

7b. The wheelchair that was pushed by the bellboy was heavier than the chainsaw.

8a. The champagne enjoyed by the diva was more expensive than the jacket.

8b. The champagne that was enjoyed by the diva was more expensive than the jacket.

9a. The sculpture remembered by the violinist was as impressive as the pottery.

9b. The sculpture that was remembered by the violinist was as impressive as the pottery.

10a. The log lifted by the body builder was not as heavy as the washing machine.

10b. The log that was lifted by the body builder was not as heavy as the washing machine.

11a. The necklace loved by the pope was kept safe with the key.

11b. The necklace that was loved by the pope was kept safe with the key.

12a. The doll embraced by the girl was next to the telephone.

12b. The doll that was embraced by the girl was next to the telephone.

High PastP verbs

13a. The CD released by the trumpeter was recorded without the harp.

13b. The CD that was released by the trumpeter was recorded without the harp.

14a. The camera selected by the ballerina was more expensive than the bicycle.

14b. The camera that was selected by the ballerina was more expensive than the bicycle.

15a. The parcel expected by the business woman was delivered without a stamp.

15b. The parcel that was expected by the business woman was delivered without a stamp.

16a. The bench painted by the gardener was next to the car.

16b. The bench that was painted by the gardener was next to the car.

17a. The song recorded by the nun was about the flower.

17b. The song that was recorded by the nun was about the flower.

18a. The cheque accepted by the plumber was signed with the quill pen.

18b. The cheque that was accepted by the plumber was signed with the quill pen.

19a. The crime scene described by the woman was shown on the TV.

19b. The crime scene that was described by the woman was shown on the TV.

20a. The gift received by the librarian was as large as the suitcase.

20b. The gift that was received by the librarian was as large as the suitcase.

21a. The bag investigated by the palace guard was full of sweets.

21b. The bag that was investigated by the palace guard was full of sweets.

22a. The jewellery protected by the janitor was inside the case.

22b. The jewellery that was protected by the janitor was inside the case.

23a. The dress designed by the bride was perfect with the shoes.

23b. The dress that was designed by the bride was perfect with the shoes.

24a. The book reviewed by the teacher was left on the desk.

24b. The book that was reviewed by the teacher was left on the desk.

Appendix 3

Table 12. A list of the verbs for each verb type used in Experiment 2. The log ratio of the past participle frequency (VVN) and the past tense frequency (VVD) calculated from our normative study and the one calculated from the BNC corpus for each verb.

Verb type	Verb	log(VVN/VVD) Completion study	log(VVN/VVD) BNC
Low PastP verbs	hunt	0.00	0.06
	entertain	-3.71	0.14
	want	1.30	-2.12
	scratch	-0.59	-1.06
	watch	1.30	-1.27
	visit	1.04	-0.85
	push	0.19	-0.39
	enjoy	0.19	-0.57
	remember	-0.59	-0.69
	lift	1.30	-0.34
	love	-0.19	-1.01
	embrace	-1.04	-0.23
High PastP verbs	release	3.71	0.94
	select	1.61	2.19
	expect	2.56	1.78
	paint	2.56	1.95
	record	2.00	1.63
	accept	1.30	1.11
	describe	0.80	1.13
	receive	2.00	0.61
	investigate	0.80	1.17
	protect	0.00	1.45
	design	3.71	2.48
	review	2.00	0.98

Footnotes

[i] In Allopenna et al. (1998), participants were well acquainted with the lexical items and the pictures of displayed referents before hearing the actual experimental instruction. In contrast, in our experimental setting, though there was a preview period (1s), participants were not informed of the lexical content prior to the experimental session. Also, it was often not possible to identify a verb at its first or second syllables as there is virtually no restriction on the set of possible verbs (e.g., *disagree* shares its first and second syllables with a transitive verb *disappoint*). Thus, even though some models of eye-movements such as EZ-Reader and SWIFT assume the latency of saccade programming to be 110-125ms (Engbert, Longtin, & Kliegl, 2002; Reichle, Rayner, & Pollatsek, 2003), it may seem too conservative to estimate it in the current experimental setting to be around 100ms. For this concern, we also performed our analysis taking the time period starting from 200 ms following the verb onset but obtained essentially the same results. It is still an empirical question how much of an influence preactivation of lexical items, explicit instruction, or time pressure can have on saccade programming latency.

[ii] A previous study showed that vowel duration as well as pitch pattern of structurally ambiguous verbs plays a role in identifying syntactic structures (Beach, 1991). Although no study, to our knowledge, has so far examined such information with intransitive verbs when it was followed by a prepositional phrase and when it was not followed by any adjunct, it is possible that the shorter duration of Intransitive+CONJ verbs may have signaled the absence of an upcoming adjunct. We thank one of the reviewers for pointing this out.

[iii] It appears from Figure 6 that there are overall more looks to the relative clause agent entity with low PastP verbs than with high PastP verbs, although the difference in the 100 – 800 ms interval was not statistically supported (no effect of Verb Type as well as no interaction between Verb Type and Time Window). One possibility for this is the difference in frequency in which the verb occurs with an explicit by-agent. It is possible that high PastP verbs may occur less often with a by-agent, resulting in less looks to the agent entity. We checked this with the British National Corpus and found that there is no difference in probabilities of the verb occurring with an explicit by-agent between the two groups of verbs (0.15 for low PastP verbs and 0.13 for high PastP verbs, $t(21.89) = 0.29$, $p = 0.77$ with an independent t-test). Given that there was no difference in plausibility for possible thematic role relations between depicted entities, shown by the plausibility rating study, it is most likely that this potential baseline difference is due to different sets of pictures used for the two groups of verbs. Note, however, what is crucial for our argument is the interaction between Verb Type and Clause Type, which provides evidence for an influence of past participle frequency information on the difficulty due to structural ambiguity.