# No prediction error cost in reading

Frisson, Steven; Harvey, David; Staub, Adrian

*DOI:*
10.1016/j.jml.2017.04.007



*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Frisson, S, Harvey, D & Staub, A 2017, 'No prediction error cost in reading: evidence from eye movements', *Journal of Memory and Language*, vol. 95, pp. 200-214. https://doi.org/10.1016/j.jml.2017.04.007

# No Prediction Error Cost in Reading: Evidence from Eye Movements

Steven Frisson

University of Birmingham, UK

David R. Harvey

University of Warwick

Adrian Staub

University of Massachusetts, Amherst

Correspondence concerning this article should be addressed to Steven Frisson, School of Psychology, University of Birmingham, Edgbaston B15 2TT, UK.

Contact: s.frisson@bham.ac.uk

tel. ++44(0)121 414 8201

fax ++44(0)121 414 4897

Word count main text: 7,464

## Abstract

Two eye movement while reading experiments address the issue of how reading of an unpredictable word is influenced by the presence of a more predictable alternative. The experiments replicate the robust effects of predictability on the probability of skipping and on early and late reading time measures. However, in both experiments, an unpredictable but plausible word was read no more slowly when another word was highly predictable (i.e. in a constraining context) than when no word was highly predictable (i.e. in a neutral context). In fact, an unpredictable word that was semantically related to the predictable alternative demonstrated facilitation in the constraining context, in relatively late eye movement measures. These results, which are consistent with Luke and Christianson's (2016) corpus study, provide the first evidence from a controlled experimental design for the absence of a prediction error cost, and for facilitation of an unpredictable but semantically related word, during normal reading. The findings support a model of lexical predictability effects in which there is broad pre-activation of potential continuations, rather than discrete predictions of specific lexical items. Importantly, pre-activation of likely continuations does not result in processing difficulty when some other word is actually encountered.


Keywords: eye movements, reading, predictability, prediction cost

When reading or listening, it is sometimes possible to anticipate which word will appear next in a sentence, and it is very clear that the predictability of a word has consequences for processing during incremental comprehension. Eye movement studies have shown that a predictable word, as measured by the word's cloze probability (i.e. the proportion of participants in an off-line production task who complete a sentence fragment using the word; Taylor, 1953), receives shorter eye fixations during reading than does an unpredictable word (e.g. Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996; Zola, 1984) and is less likely to be fixated at all, i.e. the word skipping rate is higher (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Rayner & Well, 1996). Staub (2015) reviews this literature. Evidence that predictability can facilitate processing also comes from electrophysiological data. The N400 is a negative peak in Event Related Potential (ERP) recordings that occurs approximately 400 ms after the onset of a word during either visual or auditory presentation of sentences. The amplitude of this response is increased when a word provides a poor semantic fit in its context (Kutas & Hillyard, 1980, 1983), but also when a word is relatively unexpected, as measured by cloze probability (Federmeier & Kutas, 1999; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas & Hillyard, 1984).

The main focus of the present study is on the processing of a word that is relatively *un*expected in its context. We use eyetracking during reading to address two questions regarding processing of an unexpected word. The first is whether there is an additional processing cost associated with encountering a low cloze probability word in a context in which another word is highly predictable. A word may have low cloze probability following a context for which there is no word that has high cloze probability, i.e. a *neutral* context. On the other hand, a word may have low cloze probability when there is some other word that does have high probability as a cloze continuation, i.e. in a *constraining* context. Assuming that a given low

cloze probability word is a sensible continuation in both cases, is there nonetheless a processing disadvantage when this word occurs in the constraining context? We refer to such a potential cost as a *prediction error cost*, as it would presumably reflect inhibition related to the fact that a specific lexical prediction has not been satisfied. A prediction error cost would suggest that, in a constraining context, a comprehender does specifically expect the high cloze probability continuation. The lack of a prediction error cost, on the other hand, would suggest that readers do not typically maintain specific lexical expectations that are undermined when an unexpected, but sensible, word is encountered instead. In the General Discussion, we return to the question of how such a negative answer is best interpreted.

The second question we address is whether the processing of a low cloze probability word in a constraining context is modulated by the degree to which this word is related in meaning to the more predictable alternative. If an unpredictable word is closely related in meaning to a more expected word, does this relationship facilitate processing? Again, the answer to this question has the potential to inform our understanding of the nature of contextually based expectations in language processing. A positive answer might suggest that context may generate expectations at the level of semantic features or semantic categories, not merely at the level of specific words. However, depending on the empirical details, a positive answer might also implicate the role of late, integrative processes. Specifically, a word that is related to a predictable word may be easier to integrate into the discourse context, even if the word itself is not actually expected.

Several ERP experiments (e.g. Federmeier et al., 2007), thoroughly reviewed by Van Petten and Luka (2012; see also Kutas, DeLong, & Smith, 2011), have investigated the question of whether the N400 is increased for a low-predictability word in a constraining context

compared to a neutral context. Van Petten and Luka's (2012) review concludes that while the amplitude of the N400 is very sensitive to the gradations of cloze probability, it is *not* modulated by violation of expectations; they remark that, "current data….provide little hint that amplitudes are increased when an hypothesis/expectation/prediction is disconfirmed" (p. 180). However, they also discuss several studies that have reported a distinct effect in these circumstances, a late positivity that tends to have a frontal scalp distribution. More recently, DeLong, Quante, and Kutas (2014) confirmed such a late positivity. In sum, while there is some ERP evidence that contextual constraint may modulate the electrophysiological response to a low cloze probability word, the effect appears to be distinct from the N400 effect of a word's predictability itself, and is delayed relative to this effect.

However, another ERP study using a different logic (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; see also see also DeLong, Urbach, & Kutas, 2005; Wicha, Moreno & Kutas, 2004; though cf. Nieuwland et al. 2017) arguably points to the conclusion that encountering unpredicted input in a constraining context is disruptive. Van Berkum et al. made use of the fact that Dutch prenominal adjectives agree with the following noun in syntactic gender. Two ERP experiments revealed a distinct effect when readers encountered a prenominal adjective that did not agree in gender with a high cloze probability noun, though the noun had yet to be encountered. Van Berkum et al., who also obtained a hint of a similar effect in a self-paced reading experiment, interpreted these findings as evidence that language comprehenders do predict specific words, with disruption appearing as soon as these predictions are shown to be incorrect.

ERPs have also been used to address the second of the questions outlined above, the question of whether processing of a low cloze probability word is influenced by semantic

relatedness between this word and a more expected continuation. In a well-known study, Federmeier and Kutas (1999; see also Federmeier, McLennan, De Ochoa & Kutas, 2002) manipulated whether the final noun in a constraining sentence was an expected target, an unexpected target from the same semantic category as the expected target, or an unexpected target from a different category (e.g. "The gardener really impressed his wife on Valentine's Day. To surprise her, he had secretly grown some *roses/tulips/palms"*). They found a reduced N400 for unexpected, within-category targets, compared to unexpected, between-category targets.

One potential drawback of most ERP experiments, however, is the unnatural presentation of stimuli (see Clifton & Staub, 2011; Rayner, Ashby, Pollatsek, & Reichle, 2004; Schotter, Tran, & Rayner, 2014). Words or phrases are typically presented for a fixed amount of time in Rapid Serial Visual Presentation (RSVP) format. On the one hand, the time available to process each word is usually substantially longer than the time taken to read a word in normal reading. On the other hand, re-reading is not possible. In addition, word skipping, which is a normal part of reading and can be indicative of successful pre-processing of upcoming words (more predictable words are more often skipped, e.g. Drieghe, Rayner, & Pollatsek, 2005), is precluded. As a result, it is possible that participants in ERP studies generate predictions more actively, or in a different way, than would be the case in normal reading. Evidence does indeed suggest that prediction can be context- or task-dependent (e.g. Brothers, Swaab, & Traxler, 2017; Huettig & Mani, 2016). For example, in a visual world study, Huettig and Guerra (2015) found that with a normal speech rate, anticipation effects were observable if participants had a long preview (4 seconds) of the visual scene, but not when the preview was shorter (1 second). Of direct relevance to the interpretation of predictability effects in ERP experiments is the finding by Dambacher, Dimigen, Braun, Wille, Jacobs, and Kliegl (2012) that the rate of RSVP

presentation influences the size of predictability-related N400 effects, with smaller effects emerging at rates that approximate natural reading. Also relevant is the recent finding by Brothers et al. (2017) that the predictability-related N400 effect with RSVP presentation was magnified when subjects were explicitly instructed to make lexical predictions, compared to when they simply read for comprehension. Together, these considerations suggest that it is critical to address the issue of prediction error cost in more natural reading paradigms.

Related results from a self-paced reading paradigm have been reported by Roland, Yun, Koenig, and Mauner (2012), who found that reading time was predicted not only by a word's cloze probability, but also by the word's degree of semantic relatedness to other cloze completions. Roland et al. compared reading times for target words after contexts such as "jabbed/attacked the angry lion with…". In a cloze task, the former verb elicited completions that were all pointed weapons (e.g. spear, knife) while the latter elicited some pointed weapons, but also other types of implements (e.g. rock, gun). They found that reading times for the completions were predicted by cloze probability, as expected, but that the semantic similarity of the target word to other possible completions independently affected reading times: Participants read words faster when other words that could have appeared in the same context were highly semantically related to the presented word, as measured by Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). Roland and colleagues suggested that this effect can arise either due to words that are semantically related to an expected word becoming strongly activated by means of spreading activation within the lexicon (e.g. Neely, 1977), or by independent activation of multiple words with semantic features that would satisfy the constraints set by the context.

An older study by Schwanenflugel and LaCount (1988), in which subjects made lexical decisions to words presented after sentence contexts, also reached the conclusion that semantic relatedness between a word and the expected completion influences processing, but found that this influence depended on the context's degree of constraint. In this study, highly constraining sentences (mean cloze probability of modal response = 88.2%) reduced lexical decision latency only for the most expected completion. However, lower-constraint (but still relatively constraining) sentences (mean cloze probability of modal response = 51.6%) facilitated lexical decisions for both expected and semantically related completions, but not for completions unrelated to the expected word. Schwanenflugel and LaCount suggested that a constraining context establishes restrictions on the features a possible completion must possess. When a word's semantics matches these restrictions, processing is facilitated. The more restrictions a context imposes, the less likely a word is to meet all of them, and so facilitation is less probable. According to this suggestion, a completion that matches the restrictions imposed by a context, even if the completion is not the word with the highest cloze probability, should be processed more easily.

However, most relevant to the present study is a recent eye movement in reading study by Luke and Christianson (2016). Luke and Christianson had participants read passages for which cloze values were calculated for each word in the text. In regression models, they found clear predictability effects on multiple eye movement measures, but they also directly investigated possible prediction error cost. For all words in their corpus that were not the most frequently predicted word in the corresponding cloze task, they assessed whether processing difficulty increased as the context's level of constraint increased. They found that not only was this not the case, but increased constraint seemed to facilitate processing of unpredictable words. There were more instances of word skipping and fewer refixations when an unpredictable content

word was encountered in a more constraining context, and for function words, there were also facilitative effects of constraint on reading time measures. Luke and Christianson interpreted these results in terms of the suggestion by Staub, Grant, Astheimer, and Cohen (2015) that words that are produced with low (but non-zero) cloze probability in constraining contexts are themselves highly activated.

Luke and Christianson's results also assessed whether semantic relatedness between an unpredictable word and other possible continuations influence reading time on this word. They scored semantic relatedness based on the mean LSA cosine (Landauer & Dumais, 1997) between each word in the text and all responses provided in the cloze task. For content words, this LSA score was indeed predictive of eye movements in a model that also included predictability itself. There was facilitation in both early and late measures when a word was more closely related to other possible continuations.

The present study may be seen as following up Luke and Christianson's (2016) results in a controlled experimental design. This is important for three reasons. First, Luke and Christianson's corpus was characteristic of natural text in that it contained few highly predictable words. The mean cloze probability for content words was about .13, and only 5% of content words in the corpus were highly predictable, with cloze > .67. Thus, the study had limited power to assess whether there is a prediction error cost that specifically arises in the extremely constraining contexts that have generally been employed in both eye movement and ERP experiments investigating predictability effects. Second, in the present study we are able to carefully control both the plausibility of unpredictable words that are presented in constraining contexts and the degree of semantic relationship between an unpredictable word and the corresponding predictable word. Finally, the present study has the usual virtue of a controlled

experimental design in that target words may be held constant across contexts, eliminating the potential role of lexical confounds that, in a corpus study, must be statistically controlled in a regression model.

In the first of the experiments presented here, participants' eye movements were monitored as they read target words following sentence contexts that were either highly constraining or relatively neutral. In the constraining context, one possible target word was the expected word, while the other was a low cloze, but sensible, alternative. Both words had low cloze probability in the neutral context. We assessed whether there is a reading time cost associated with encountering an unpredictable word in a constraining context, compared to the same word in a neutral context. In the second experiment, we also manipulated the degree of semantic relatedness between the low cloze probability words and the high cloze probability alternative. To our knowledge, this is the first attempt to systematically investigate prediction error cost and facilitation of unpredicted semantically related words in a standard reading task.

## Experiment 1

### *Method*

*Participants***.** Forty-four undergraduate students from the University of Birmingham participated in the eye-tracking experiment for course credit or remuneration. All were native British English speakers and had normal or corrected to normal vision.

*Materials***.** We selected 44 item quartets (see Appendix A and Table 1) after cloze testing in which students were given a sentence fragment up to the target word and asked to complete it with the first word or words that came to mind. These participants did not take part in the eye movement study. The sentence fragments for the selected items were completed by 26

participants on average (range: 17-30). Mean cloze probability was 70.2% for the Constraining context with Predictable word (CP), 1.8% for the Constraining context with Unpredictable word (CU), 0.5% for the Neutral context with the Predictable word from the CP condition (NP), and 1.2% for the Neutral context with the Unpredictable word from the CU condition (NU). Paired-samples *t*-tests showed that the cloze values for CP were significantly higher than for both CU and NP (*p*s > .001). Most importantly for testing the prediction error cost, the cloze values for CU and NU were comparable ($t < 1$). The predictable and unpredictable target words were matched for length (4.8 characters on average for both) and CELEX frequency ($t < 1$, using N-Watch, Davis, 2005).

The sentence frames for the predictable and unpredictable target words were identical. Note that the classification of (un)predictable target word is based on the answers for the constraining contexts; all target words in the neutral contexts had a low cloze value. We also added a second sentence in order to make the items similar in length to items from other experiments with which these items were intermixed, but eye movement data for this second sentence were never analyzed.

We also checked whether the sentences were seen as relatively *plausible*, to ensure that any differences between conditions could not be attributed to differences in the plausibility of the target word as a continuation. A different group of 18 participants used a 7-point scale to judge the plausibility of the sentence fragments up to and including the target word. We included an equal number of items which we considered to be implausible, in order to encourage participants to use the whole scale. Mean plausibility ratings for the four conditions were: CP: 6.5, CU: 5.7, NP: 5.8, NU: 5.9, and implausible filler items: 1.9. While all conditions were seen as highly plausible, the ratings were significantly higher for CP than for all other

conditions (all *p*s < .001), while the other conditions did not significantly differ from each other (all *p*s > .27). We believe that the somewhat higher value for the CP condition is likely due to subjects' difficulty in rating the stimuli based on plausibility rather than predictability itself.[1]

*Procedure***.** Participants were run individually on an Eyelink 1000 SR Research eyetracker with tower mount. Participants were instructed to read for comprehension. In order to keep their attention, yes/no questions were asked after 36% of the sentences (accuracy averaged 93.2%). Viewing was binocular, but data were only collected for the right eye.

Before each item, a drift check was performed in the middle of the screen. Then a black rectangle was shown just to the left of the first character of the first sentence; the participant's fixation on this rectangle triggered the presentation of the sentence. Participants read the items silently and terminated each item by pressing a button on a buttonbox. If a question followed, this was answered with one of two buttons on the buttonbox. The experimenter calibrated the participants' eye position at the beginning of the experiment and whenever necessary, as indicated by the drift check. The entire experiment lasted about 30 minutes.

Stimuli were divided over four lists, with an equal number of items for each condition per list. Participants were randomly assigned to one of the lists and presentation of the items was pseudo-randomized so that no two items from the same condition were presented consecutively.

*Results*

[1] Indeed, the predictability/plausibility correlation was relatively high ($r$ = .38, $p$ < .001). Including plausibility in the statistical models we report below did not result in meaningful changes in the pattern of results.

Trials with track loss and trials that contained a blink in the region of interest were removed (2.5%). Short fixations (<80 ms) within 1 character position of another preceding or following fixation were automatically combined. Other fixations less than 80 ms in duration were removed, as were fixations greater than 1500 ms in duration. For the gaze duration measure, reading times over 750 ms were excluded (3 observations).

We constructed linear mixed models of the eye movement data, using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.2.2. (R Core Team, 2015). For binomial data (regressions, skips), we carried out logistic regressions in a generalized linear mixed effect model. The following measures were analysed: skipping percentage, first fixation duration (the duration of the first fixation on a region on first-pass), single fixation duration (reading time on a word or region if only one fixation occurred on first-pass), gaze duration (sum of all first pass fixations on a region before looking elsewhere), percentage of first-pass regressions out of a region, regression-path duration (sum of all fixations from first entering a region to going past that region; this measure can include refixations on previous text), and total time (sum of all fixations on a region). The region of interest was the target word itself, excluding any determiners or adjectives.

We are interested in three specific comparisons: (1) constraining context – unpredictable target word (CU) vs. constraining context – predictable target word (CP), (2) neutral context – predictable target word (NP) vs. constraining context – predictable target word (CP), and (3) constraining context – unpredictable target word (CU) vs. neutral context – unpredictable target word (NU). The first two comparisons test the typical predictability effect and evaluate whether both our choice of target word (1) and context (2) were appropriate. The crucial comparison is (3), as it tests prediction error effects: If there is a cost for having predicted

the wrong word, or having a different word highly activated, then CU should be more difficult to process than NU.

The three comparisons were tested in a single model for each eye movement measure. (We note that separate models independently testing each of the three comparisons deliver identical patterns of statistical significance to those reported below.) Because the contrasts are not orthogonal, the coefficient matrix used in the model was the transposed generalized inverse of a contrast matrix defined by the three pairwise comparisons of interest (Venables & Ripley, 2002). All models of reading times included the maximal random effects structure, i.e., including subject and item intercepts, and subject and item slopes for each of the three contrasts. For the logistic models of skipping and regressions out, random slopes were eliminated to obtain convergence.

Table 2 presents the means and standard errors, by subject, for each eye movement measure. Table 3 provides the results of the statistical analysis. The results reveal predictability effects in all eye movement measures. Facilitation for a predictable word was found both when this word is compared to an unpredictable word in the same constraining context (CU vs. CP), and when this word is compared to the same target word in a neutral context (NP vs. CP). All effects reached significance (defined by $|t|$ or $|z| > 2$), with the exception of the skipping and regressions out measures for the CU vs. CP comparison. On the other hand, there was no hint of a prediction error cost: An unpredictable word in a constraining context never resulted in a processing cost compared to when that word appeared in a neutral context (CU vs. NU). In fact, in all measures the numerical trend was in the opposite direction, i.e., facilitiated processing in the CU condition compared to the NU condition. In the first-pass regressions measure, this reversed effect actually reached signficance.

## Experiment 2

In Experiment 1, there was no indication of a prediction error cost; processing of the unpredictable word was not more difficult in the constraining context compared to the neutral context. Indeed, there were fewer regressions from this word in the constraining context, and reading time measures showed a trend in the same direction. However, we did not manipulate the semantic relationship between the predictable and unpredictable target words. It is possible that facilitation for an unpredictable word in a constraining context arises only when this word is semantically related to the predictable target. Experiment 2 addresses this issue.

### *Method*

*Participants***.** Forty-eight undergraduate students from the University of Birmingham took part in the eye-tracking experiment.

*Materials***.** After successive rounds of cloze tests, we selected 60 constraining and 60 neutral sentence fragments. These were paired with the predictable word from the constraining fragments, a semantically related word, and a semantically unrelated word (see Appendix B for all stimuli). Hence, each item consisted of 6 versions: Constraining context with Predictable target word (CP), Constraining context with an unpredictable but semantically Related target word (CR), Constraining context with an unpredictable and semantically Unrelated target word (CU), and their Neutral context counterparts (NP, NR, NU). Predictability, as assessed in a cloze task (with a minimum of 25 responses per sentence fragment), was higher for the CP condition than all other conditions (CP: 76.6%, CR: 0.8%, CU: 0.5%, NP: 0.3%, NR: 0.3%, NU: 0.6%). Crucially, the predictability of the unpredictable word targets was comparable in the constraining and neutral conditions (for related target words: $p > .08$; for unrelated target

words: $p > .69$). In addition, the predictability of the semantically related and unrelated target words was comparable in both the constraining ($p > .45$) and neutral ($p > .20$) conditions.

We checked semantic relatedness of the target words in two different ways. First, fifty-five participants, divided over 3 lists, rated the relatedness of pairs of words on a 7-point scale, with 1 being not at all related and 7 very closely related. Each trial compared two of the target words in a given item, and each word pair was rated by at least 18 participants. The relatedness score of the predictable and the semantically related target words (5.4) was significantly higher than the predictable-unrelated score (2.5, $p < .001$) and the related-unrelated score (2.2, $p < .001$). The last two scores did not differ significantly from each other ($p > .08$). Second, we calculated the Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) scores for each word pair. Again, the predictable-related scores (.36) were significantly higher, indicating greater semantic overlap, than the predictable-unrelated (.13; $p < .001$) and the related-unrelated scores (.12; $p < .001$), which did not differ from each other ($p > .28$). The target words were of the same length in all conditions (5.2 characters on average). Due to selection restrictions (semantically related words of the same length), the frequency (per million) of the predictable words (75.3) was somewhat higher than the related words (38.6, $p < .05$), but did not differ significantly from the unrelated words (56.8, $p > .25$). More importantly, the frequency difference between the related and unrelated target words was not significant ($p > .13$) and, if anything, the numerically lower frequency of the related target words compared to the unrelated target words would work against any facilitation from semantic relatedness.

Finally, we carried out a plausibility check for the items used in the experiment (see Experiment 1 for details; the items of both experiments were combined in the same test). Mean plausibility scores were high overall: CP: 6.6, CR: 5.8, CU: 4.8, NP: 5.9, NR: 5.7, NU: 5.9. As

before, we suspect that the higher score for the predictable target (*p*s < .001, compared to all other conditions) is at least partially due to participants judging predictability rather than the plausibility. The plausibility mean for the CU condition was also lower than in the other conditions (all *p*s < .001). If this difference were to have effects in the eye movement record (and we note that all existing evidence of plausibility effects comes from comparisons involving words that are very implausible, e.g. Staub, Rayner, Pollatsek, Hyona & Majewski, 2007, unlike the current case) it would result in longer reading times in the CU condition than the NU condition. To anticipate the results, there was no indication of such an effect, suggesting that this difference in plausibility was too small to be functionally relevant.

*Procedure***.** The procedure was the same as for Experiment 1, with 27% of the sentences being followed by a yes/no comprehension question (mean accuracy 96.5%). Participants saw two conditions of each item quartet, one with a constraining and one with a neutral preceding context, but with a different target word.

*Results*

Data handling was identical to Experiment 1. Less than 1% of the data was removed due to track loss or blinks in the region of interest. We tested the following comparisons: (1) predictable word in a constraining context vs. all other conditions, which tests the predictability effect; (2) unpredictable word that is semantically related to the predictable word, in a constraining vs. neutral context (CR vs. NR); (3) unpredictable word that is not semantically related to the predictable word, in a constraining vs. neutral context (CU vs. NU). We tested these three orthogonal contrasts in a single mixed-effects model for each eye movement measure by coding the relevant conditions as .5 or -.5. All models of reading times used the maximal random effects structure (i.e. random subject and item intercepts, and random subject

and item slopes for each of the three contrasts). For the logistic models, the random slopes were removed for convergence.

Table 4 presents the averages for each eye movement measure, and Table 5 provides the results of the statistical models. The results are easily summarised: (1) clear and significant predictability effects were observed in all measures, with the CP condition showing facilitation relative to the other conditions; (2) there was no evidence of a prediction error cost, for either the semantically related or unrelated conditions, as the CR and CU conditions did not have slower reading times, fewer skips, or more regressions than the corresponding NR and NU conditions; and (3) there was facilitation if the target word is unpredictable but semantically related to a predictable word, evident in an advantage for the CR condition compared to the NR condition. The related word in a constraining context elicited significantly fewer regressions than in a neutral context, and total time was also shorter in the constraining context. Non-significant numerical trends in the same direction were present in all other measures. These patterns were not found for the unrelated word.

## General Discussion

When readers encounter a predictable word, processing of that word is facilitated, as indicated by more skipping, shorter fixation durations, and fewer regressions. We replicated this classic predictability effect in both experiments and across a wide range of measures. This effect was found both when the predictable target word was compared to an unpredictable word in the same constraining context, and when compared to the identical word in a neutral context. These results are consistent with the already strong evidence (Staub, 2015) that a predictable word is pre-activated sufficiently to affect even the earliest reading measures.

Our second finding is more consequential as it sets limits on the type of model that can explain predictability effects. In neither experiment did we observe any evidence of a prediction error cost, i.e. reading was not slowed when an unpredictable target word appeared in a sentence in which a different word was highly predictable. In fact, it appears that processing may be somewhat facilitated in this instance, a finding that we discuss further below. The lack of prediction error cost is important, as a model that assumes that one specific word is predicted and the activation of other words is suppressed, or a model that assumes difficulty when the input is inconsistent with a highly pre-activated lexical item, would not seem able to account for the present data.

It is important to note that the lack of a significant prediction error cost in these experiments is probably not due to insufficient power. The experiments had sufficient power to detect the expected predictability effects in both early (e.g., skipping rate, first fixation duration) and later (e.g., gaze duration, regression path duration) measures. Moreover, these effects were evident when assessed based on comparison of the same word in two contexts, or based on comparison of different words in the same context. But in most measures in both experiments, the trend in the comparison of the unpredictable target in constraining and neutral targets was actually in the direction of faster processing for this word when it was encountered in a constraining context. In some measures this benefit, as opposed to cost, was actually significant, specifically for unpredictable words that were semantically related to the predictable alternative. Finally, the evidentiary value of the null effects reported here is increased due to the fact that they may be regarded as confirming, with a more direct test, the lack of prediction error cost (indeed, benefit in some circumstances) obtained in other eye movement experiments, as we discuss below.

The lack of a prediction error cost is in line with findings from other eye movement experiments. Frisson, Rayner, and Pickering (2005), investigating the relationship between transitional probability and predictability, did not observe a prediction error cost either, though the two reported experiments were not designed to specifically look at this issue. In addition, the cloze values for the predictable word were much lower (approximately 20%) than in the present experiments (70% and 77%, respectively). Hence, if the predictable word is only moderately predictable, then possible competition effects might be too small to be picked up. Similarly, Luke and Christianson (2016) found no evidence of a prediction error cost in their eye movement corpus. While the average predictability in Luke and Christianson's (2016) corpus study was very low (13% for content words, 22% for function words), the predictability of the predictable word in the current experiments was very high, so the absence of a cost when some other word is encountered may be regarded as more surprising.

In sum, there is now clear evidence against the idea that encountering an unpredictable word in a context in which some other word is highly predictable results in a measurable cost, in terms of a slowdown in reading. It is important to note that this result does not directly argue against models that propose learning of language statistics and language structure based on prediction error (e.g. Christiansen & Chater, 2016; Elman, 1990). For example, in their Now-or-Never Bottleneck model, Christiansen and Chater (2016) propose an error-driven learning mechanism in which predictions are compared with the actual input. When this results in an error signal, the model is adapted in order to reduce the error. The present results allow that a neural prediction error signal may be generated when the input is unexpected, and that such an error signal may be used to update representations of language at multiple levels. The present results do show, however, that any such error signal does not slow down or interrupt current processing.

The finding that prediction error costs do not show up in on-line reading deviates from the behavioral data that have shown a cost (Forster, 1981; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985), though these experiments (naming, lexical decision) did not test natural reading. The lack of a prediction cost also contrasts with some recent neurophysiological data (e.g. Federmeier et al., 2007; see also DeLong, Troyer & Kutas, 2014), which showed an effect that may be interpreted as a delayed cost – a late, frontally distributed, positivity. Since we did not find an inhibitory effect in any measure, we suspect that the cost might be related to the idiosyncrasies of electrophysiological experimentation. For example, the presentation rate allows participants to spend more time processing each word than they would during normal reading, which could lead both to more conscious, controlled prediction strategies and to more evaluative processing of target words. The observation that the cost seems to be delayed in the EEG signal fits with this view. On the other hand, the late positivity may be interpreted as a reflection of the error signal itself, rather than as a processing cost that results from the error signal. As we note above, we see no reason to deny that there may be such a signal, even if it has no immediate processing consequence.

Our final finding is that when an unpredictable word is related to a highly predictable alternative, there is processing facilitation. Given that this effect reaches significance in only a limited set of relatively late eye movement measures (though the numerical trend is present in all measures), we would regard this result with some scepticism, were it not for the fact that the result corroborates existing behavioral (e.g. Schwanenflugel & LaCount, 1988; Roland et al., 2012) and electrophysiological data (e.g. Federmeier & Kutas, 1999), and indeed, confirms a result from a corpus study of eye movements in reading (Luke & Christianson, 2016;). However, the present data are the first evidence of this effect from a natural reading task in which semantic relatedness was explicitly manipulated.

Unlike Luke and Christianson (2016), who found such facilitation for almost all measures in their corpus study, we found significant facilitation only in intermediate (first-pass regressions) and late (total time) reading measures. It is possible that this difference is more apparent than real; in every measure in the present Experiment 2, the CR condition showed facilitation compared to the NR condition, and a reasonable inference from this pattern is that in fact there is general facilitation, though the effect is small and does not always reach significance. As has often been noted (e.g. Francis, 2013), a trend in the same direction as a previously obtained result, even when $p > .05$, should be regarded as increasing, rather than decreasing, our confidence in the original result. If, on the other hand, the differences in patterns of significance across the two studies are taken seriously, it may be relevant that our study and Luke and Christianson's computed semantic relatedness in somewhat different ways. Luke and Christianson computed the mean LSA score between all cloze responses and the target word, whereas we focused on the relatedness between the target and a specific, highly predictable, alternative. At present, we cannot say to what extent this methodological choice matters.

The fact that there is facilitated processing of an unpredictable word that is semantically related to a highly predictable alternative may suggest some kind of partial pre-activation of this semantically related word. This pre-activation could be due to spreading activation starting from the pre-activation of the predictable word, or it could be due to the context independently activating a set of words with certain semantic features (see Roland et al., 2012; Schwanenflugel & LaCount, 1988). It should be noted, however, that the constraining context did not make the unrelated words implausible, hence, the contextual constraints could not have been so strong as to exclude these continuations. Indeed, there was no sign of processing cost for the unrelated target word in the constraining context compared to the neutral context. Thus, it appears that

context may provide multiple words with a boost in activation, while the activation levels of the other words are not affected. In general, contexts are hardly ever so constraining that only one or a few words[2] can plausibly be used as continuation, making it unlikely that these constraints are being used as *ex*clusion criteria. Instead, it is possible to envision these constraints as *in*clusion criteria, with words matching those criteria or features becoming more available. A word's predictability is then the level of pre-activation based on both extrinsic (contextual, which can also include, for example, visual context; for a discussion of prediction in the visual world paradigm, see e.g. Altmann & Kamide, 2007; Altmann & Mirković, 2009) and intrinsic factors (e.g. word frequency, or the base level of a certain word occurring; see Staub, 2015, for a discussion of the additive rather than interactive nature of some these factors).

On the other hand, there is a clear problem for an account that attributes facilitation of an unpredictable, semantically related word to pre-activation of this word, if the specific pattern of significance across the eye movement measures is taken at face value. We did not obtain significant facilitation for this word in the early measures (skipping, first fixation duration, gaze duration) that are most reliably influenced by predictability itself. Instead, we found that readers were relatively unlikely to regress from a target word that was semantically related to the predictable alternative, and had reduced reading time on the total time measure, i.e. when re- reading is taken into account. These details suggest a different account. It is possible that a word that is semantically related to a predictable target is more easily integrated into the discourse context, resulting in a reduction of whatever difficulty may otherwise arise. Consider the example in Table 1. When the word *garden* appears instead of the highly predictable word *church*, the reader is required to modify her or his event representation from one in which the

[2] However, this might be more likely when the next word is a function word.

priest is trying to increase attendance at a service, to one, perhaps, in which the priest is trying to get more parishioners to come outside after the service, or to come to some other kind of event in the church's garden. But when the word *sermon* appears instead of *church*, no such change of event representation is required. Though this will vary from item to item, it may generally be the case in these experimental materials, or more generally, that encountering a word that is semantically unrelated to a predictable word requires such a revision, while encountering a related word does not.

At present, we are agnostic between a pre-activation account of facilitation for the semantically related word and one that emphasizes such late integrative processes, given the uncertainty about the specific eye movement measures in which there is an advantage for the semantically related word in the constraining context. The specific pattern of significant effects in the eye movement record is most consistent with the second account: Semantic relatedness between the presented word and the predictable word seems to reduce difficulty, rather than facilitating processing (see Reichle, Warren, & McConnell, 2009, for an implemented model demonstrating how regressive eye movements may result from post-lexical integration difficulty). However, additional research is certainly needed to clarify this pattern. We also point out that experiments using the boundary paradigm in conjunction with manipulations of predictability and semantic relatedness (e.g. Veldre & Andrews, *in press*) may help to arbitrate between these accounts.

Yet another interpretation of the facilitation that we observed for words that were semantically related to a predictable word relates to the idea of “Good Enough” language processing (e.g. Ferreira & Patson, 2007). It is possible that on occasion, possibly due to the fact that not all of the comprehension questions required in-depth processing, readers processed

the related word only shallowly and didn't notice that rather than the predicted word, a semantically related word was presented. This situation would be somewhat akin to the processes involved in the Moses illusion (Erickson & Mattson, 1981). As a result, participants might be less likely to go back to reread part of the text. It should be noted, however, that there is little evidence for contextually-driven misreading of words, except in cases of high-frequency orthographic neighbors (Gregg & Inhoff, 2016; Slattery, 2009). In addition, recent research has shown that comprehension question difficulty modulates primarily the probability of re-reading from near the end of sentences, rather than first-pass reading behavior (Weiss, Kretzschmar, Shlesewsky, Bornkessel-Shlesewsky, & Staub, in press).

In general, we believe that the present results are best accommodated by a model in which predictability effects arise by means of pre-activation of multiple lexical items, roughly in proportion to their predictability in the cloze task, rather than by means of discrete prediction of a specific word (see Staub, 2015; Staub et al., 2015). This pre-activation results in facilitated processing of a predictable word, but with no discernible cost for an unpredictable word, as long as this word is also plausible. Arguably, such a mechanism is well-suited for the demands of actual language processing, where words are rarely highly predictable (Gough, 1983; Luke & Christianson, 2016), so that encountering an unpredictable word is the rule rather than the exception. We note that the Surprisal (Levy, 2008) computational metric also predicts this pattern. Processing difficulty under Surprisal is a function only of a word's conditional probability in its context, regardless of sentence constraint, i.e. the Surprisal of an unpredictable word does not depend on whether one other word is highly probable, or whether there are many possible continuations with low probability. The view that predictability effects arise by means of graded pre-activation of multiple lexical items explains why prediction error costs are not observed, why it is possible to find some degree of facilitation for unpredicted words in a

constraining context, and why this facilitation is likely to be more substantial when the unpredicted word comes from the same semantic field as the predicted word. It also accounts for the finding that gradations in cloze affect the N400 (see Van Petten & Luka, 2012), and the results of Luke and Christianson (2016), who found that predictability effects are already apparent at the low end of the predictability scale.

In conclusion, the present experiments represent the first attempt to systematically investigate prediction error cost during normal reading, in a controlled experimental design with highly predictable target words. We observed no evidence for prediction error cost, whether an unpredictable word was semantically related or unrelated to the corresponding predictable word. Indeed, we found some evidence of facilitation for semantically related unpredictable words, with the strongest effect found in measures taking into account regressive re-reading. These results provide support for the view of predictability effects according to which they arise due to graded pre-activation, as opposed to discrete prediction. Reading is facilitated by pre-activating words that are likely to be upcoming, while words that are unlikely to follow are not de-activated or inhibited.

Acknowledgements

The authors thank Denis Drieghe for his help with stimuli creation and analyses, and Bernhard Angele and Andrew Olson for their help with the data analyses.

References

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24*(4), 477-492.

Altmann, G. T. M., & 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*(4), 502-518.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364-390.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language, 93*, 203-216.

Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition, 14,* 509-522.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62.

Clifton, C., Jr., & Staub, A. (2011). Syntactic influences on eye movements in reading. In: S. P. Liversedge, I. D. Gilchrist and S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 895-909). Oxford, UK: Oxford University Press.

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia, 50(8),* 1852-1870.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70.

Delong, K. A., Quante L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61,* 150-162.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass, 8*(12), 631-645.

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203-1207.

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

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 954-969.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540-551.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133-146.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75-84.

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

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology, 33*(4), 465-495.

Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, *57*, 153-169.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 862-877.

Gough, P. B. (1983). Context, form, and interaction. In K. Rayner (Ed.), *Eye movements in reading* (pp. 203-211). New York: Academic Press.

Gregg, J., & Inhoff, A. W. (2016). Misperception of orthographic neighbors during silent and oral reading. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 799-820.

Huettig, F., & Guerra, E. (2015). Testing the limits of prediction in language processing: Prediction occurs but far from always. *21st Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2015)*, Valetta, Malta, 2015.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not.

*Language, Cognition and Neuroscience, 31*(1), 19-31.

Kutas, M., DeLong, K.A., & Smith, N.J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.) *Predictions in the brain: Using our past to generate a future* (pp. 190-207). Oxford: Oxford University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Mézière, D. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *bioRxiv*, 111807.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 720-732.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*(4), 504-509.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ-Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulleting & Review, 16,* 1-21.

Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267-279.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60-94.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science, 25*(6), 1218-1226.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 344-354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*(2), 232-252.

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1969-1975.

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

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1-17.

Staub, A., Rayner, K., Pollatsek, A., Hyona, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1162-1169.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30*, 415-433.

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*(3), 443-467.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Veldre, A., & Andrews, S. (in press). Parafoveal preview effects depend on both preview plausibility and target predictability. *The Quarterly Journal of Experimental Psychology*.

Weiss, A. F., Kretzschmar, F., Sclesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (in press). Comprehension demands modulate re-reading, but not first pass reading behaviour. *Quarterly Journal of Experimental Psychology.*

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*(7), 1272-1288.

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics,*

*36*(3), 277-284.

Table 1. Sample items in Experiments 1 and 2. Predictable word refers to the word with the highest cloze value in the constraining context. The target word is underlined.

| Experiment | Condition | Example |
|---|---|---|
| 1 | Constraining context – predictable word (CP) | The young nervous paratrooper jumped out of the plane when he heard the shots. |
| 1 | Constraining context – unpredictable word (CU) | The young nervous paratrooper jumped out of the chair when he heard the shots. |
| 1 | Neutral context – predictable word (NP) | The tired movie maker was sleeping in the plane when he was woken up by a scream. |
| 1 | Neutral context – unpredictable word (NU) | The tired movie maker was sleeping in the chair when he was woken up by a scream. |
| 2 | Constraining context – predictable word (CP) | The priest wondered how he could get more people to come to the church even though it was raining. |
| 2 | Constraining context – semantically related word (CR) | The priest wondered how he could get more people to come to the sermon even though it was raining. |
| 2 | Constraining context – semantically unrelated word (CU) | The priest wondered how he could get more people to come to the garden even though it was raining. |
| 2 | Neutral context – predictable word (NP) | The widow thought that it was a lovely church even though it was cold. |
| 2 | Neutral context – semantically related word (NR) | The widow thought that it was a lovely sermon even though it was cold. |
| 2 | Neutral context – semantically unrelated word (NU) | The widow thought that it was a lovely garden even though it was cold. |

.

Table 2. Experiment 1 mean eye movement measures (standard error), by subject. Reading times in ms, skipping and regressions in percentages.

| | Constraining context – predictable word (CP) | Constraining context – unpredictable word (CU) | Neutral context – predictable word (NP) | Neutral context – unpredictable word (NU) |
|---|---|---|---|---|
| Skipping % | 29.0 (2.8) | 24.1 (2.2) | 21.2 (2.0) | 20.8 (2.3) |
| First Fixation Duration | 206 (4.0) | 220 (5.0) | 220 (4.2) | 224 (4.5) |
| Single Fixation Duration | 205 (4.2) | 221 (5.0) | 220 (4.6) | 226 (4.6) |
| Gaze Duration | 217 (4.9) | 234 (5.8) | 232 (4.9) | 237 (5.0) |
| First Pass Regressions % | 7.8 (1.6) | 11.0 (1.9) | 14.0 (2.0) | 15.9 (2.5) |
| Regression Path | 240 (7.8) | 287 (14.6) | 282 (8.3) | 293 (10.3) |
| Total Time | 238 (9.2) | 282 (10.7) | 279 (9.2) | 291 (10.3) |

Table 3. Experiment 1 mixed effects model results. The predictability effect for different target words in the same context is tested with the CU-CP comparison, the predictability effect for the same target word in a constraining vs. neutral context is tested with the NP-CP comparison, and the prediction error cost is tested with the CU-NU comparison.

| Measure | Comparison | Estimate | SE | *t/z*-value |
|---|---|---|---|---|
| Skipping % | CU - CP | -.27 | .10 | -1.77 |
| | NP - CP | -.45 | .15 | -2.87 |
| | CU – NU | .22 | .16 | 1.35 |
| First Fixation Duration | CU - CP | 13.75 | 5.36 | 2.56 |
| | NP - CP | 12.88 | 5.15 | 2.50 |
| | CU – NU | -3.42 | 6.26 | -0.55 |
| Single Fixation Duration | CU - CP | 16.44 | 5.42 | 3.03 |
| | NP - CP | 13.97 | 5.44 | 2.57 |
| | CU – NU | -3.83 | 6.47 | -0.59 |
| Gaze | CU - CP | 17.64 | 6.42 | 2.75 |
| | NP - CP | 14.17 | 6.55 | 2.16 |
| | CU – NU | -1.35 | 6.64 | -0.20 |
| First-Pass Regressions % | CU - CP | .41 | .27 | 1.52 |
| | NP - CP | .67 | .26 | 2.57 |
| | CU – NU | -.46 | .23 | -2.03 |
| Regression Path | CU - CP | 47.42 | 17.34 | 2.73 |
| | NP - CP | 38.98 | 13.91 | 2.80 |
| | CU – NU | -3.00 | 20.68 | -0.15 |
| Total Time | CU - CP | 43.10 | 10.34 | 4.17 |
| | NP - CP | 37.24 | 10.36 | 3.59 |
| | CU – NU | -9.47 | 12.78 | -0.74 |

Table 4. Experiment 2 mean eye movement measures (standard error), by subject. Reading times in ms, skipping and regressions in percentages.

| | Constraining context – predictable word (CP) | Constraining context – related word (CR) | Constraining context – unrelated word (CU) | Neutral context – predictable word (NP) | Neutral context – related word (NR) | Neutral context – unrelated word (NU) |
|---|---|---|---|---|---|---|
| Skipping % | 23.3 (2.2) | 19.4 (1.9) | 18.1 (1.8) | 18.3 (1.6) | 16.6 (2.0) | 18.9 (1.7) |
| First Fixation | 197 (3.8) | 219 (4.7) | 215 (4.9) | 213 (3.6) | 225 (5.5) | 215 (4.3) |
| Single Fix | 197 (3.9) | 222 (5.3) | 217 (5.2) | 214 (4.0) | 228 (5.9) | 215 (4.5) |
| Gaze | 209 (4.7) | 243 (5.9) | 237 (6.4) | 233 (5.4) | 249 (6.8) | 239 (6.0) |
| First Pass Regressions % | 14.6 (2.1) | 15.1 (2.0) | 20.0 (2.4) | 20.0 (2.3) | 20.6 (2.5) | 19.9 (2.2) |
| Regression Path | 262 (11.8) | 297 (11.3) | 322 (17.5) | 308 (13.5) | 318 (12.8) | 316 (11.7) |
| Total Time | 262 (12.7) | 315 (14.8) | 341 (15.8) | 325 (15.2) | 366 (17.5) | 353 (17.4) |

Table 5. Experiment 2 mixed effects model results. The predictability effect is tested with a comparison of the predictable word in the constraining context with all other conditions. The prediction error cost is tested with the CR – NR and the CU – NU comparisons.

| Measure | Comparison | Estimate | SE | *t/z*-value |
|---|---|---|---|---|
| Skipping % | Pred Effect | -.32 | .09 | -3.70 |
| | CR - NR | .22 | .13 | 1.79 |
| | CU – NU | -.07 | .12 | -0.60 |
| First Fixation Duration | Pred Effect | 20.48 | 3.31 | 6.20 |
| | CR - NR | -6.41 | 4.61 | -1.39 |
| | CU – NU | 1.32 | 3.93 | 0.34 |
| Single Fixation Duration | Pred Effect | 22.10 | 3.41 | 6.49 |
| | CR - NR | -7.62 | 5.30 | -1.44 |
| | CU – NU | 3.13 | 4.32 | 0.72 |
| Gaze | Pred Effect | 31.27 | 4.54 | 6.88 |
| | CR - NR | -7.69 | 5.623 | -1.37 |
| | CU – NU | 0.03 | 5.42 | 0.00 |
| First-Pass Regressions % | Pred Effect | .41 | .12 | 3.39 |
| | CR - NR | -.46 | .14 | -3.36 |
| | CU – NU | .01 | .14 | .07 |
| Regression Path | Pred Effect | 50.26 | 11.45 | 4.39 |
| | CR - NR | -24.04 | 12.33 | -1.95 |
| | CU – NU | 7.67 | 17.07 | 0.45 |
| Total Time | Pred Effect | 75.25 | 9.15 | 8.22 |
| | CR - NR | -51.41 | 10.86 | -4.73 |
| | CU – NU | -10.60 | 10.10 | -1.05 |

Appendix A

Experiment 1 items. For each item, the constraining context appears first (a/b). The word before the / symbol is the predictable target word, the word following the / symbol is the unpredictable word. The same target words appeared in the neutral context (c/d). The numbers in parentheses refer to the cloze percentage for the predictable/unpredictable words, respectively.

1

(a/b) The doctor told Fred that his drinking would damage his liver/heart very quickly. Fred is not feeling too well today. (77.8/3.7)

(c/d) Yesterday Fred told his friend that they will look at his liver/heart very thoroughly. Fred is not feeling too well today. (0/3.3)

2

(a/b) The banker loaned the businessman some more money/tools for his new project. The plan was to get started within a week at most. (100/0)
(c/d) The neighbour across the road wanted to get money/tools for his building work. The plan was to get started within a week at most. (0/0)

3

(a/b) For Halloween, Liz dressed up as an ugly old witch/ghost and then went to the party. She attracted quite a lot of attention. (70.4/0)
(c/d) For her arts class, the little girl drew a witch/ghost and then showed it to everyone. She attracted quite a lot of attention. (0/0)

4

(a/b) The baby laughs and giggles when she shakes her new rattle/bottle that was bought for her by her father. It cost more than you would think. (63.0/0)
(c/d) Last weekend, good friends of ours bought us a new rattle/bottle that was meant for our newborn. It cost more than you would think. (0/0)

5

(a/b) The skilled gardener went outside to pull up the weeds/roses along the driveway. To be honest, all the plants and flowers were in disarray. (55.6/0)
(c/d) The distracted professor didn't even notice the weeds/roses along the path. To be honest, all the plants and flowers were in disarray. (0/0)

6

(a/b) He is a grouch in the morning until he has had his coffee/shower and read the newspaper. He likes to stick to his daily routine. (44.4/0)
(c/d) Carl told me that he wanted some time to have a coffee/shower and read the sports pages. He likes to stick to his daily routine. (3.3/0)

7

(a/b) Late at night, the loud crying of the hungry baby/girl got the neighbours upset. It was my sister who told me this. (55.5/3.7)
(c/d) Yesterday, the thing that was said about the baby/girl got the parents annoyed. It was my sister who told me this. (0/10.0)

8
(a/b) The child was forced to cut his juicy steak with a knife/spoon at the restaurant. It was a place where they serve French food. (81.5/11.1)
(c/d) It was strange to see that my youngest was handed a knife/spoon at the bistro. It was a place where they serve French food. (0/0)

9
(a/b) The barber gives better shaves since he was given a new sharp razor/blade as a present. His grandmother thought it was a great idea. (55.6/22.2)
(c/d) For Father's day, the son wanted to give his father a new razor/blade as a surprise. His grandmother thought it was a great idea. (0/0)

10
(a/b) I was lucky to be able to visit my favourite aunt and uncle/niece over the Christmas break. The weather had been horrible for three days already. (85.2/0)
(c/d) My younger brother didn't say much when he noticed his uncle/niece over the Easter break. The weather had been horrible for three days already. (0/0)

11
(a/b) The dog buried his bone/food under the rose-bushes in Joe's backyard. It left it there for only a short while. (88.9/0)
(c/d) The animal saw the bone/food under the table in the dining room. It left it there for only a short while. (0/26.7)

12
(a/b) The lightning from the storm struck a large limb of a tree/bush in our backyard. My wife had wanted me to cut it down for a long time. (53.8/0)
(c/d) The young amateur took a long time photographing the tree/bush in our garden. My wife had wanted me to cut it down for a long time. (3.3/0)

13
(a/b) The maid was told to wash and wax the kitchen floor/table first thing this morning. Afterwards she had to attend to the bathroom. (81.5/0)
(c/d) The tired young woman was informed that the floor/table first needed waxing this morning. Afterwards she had to attend to the bathroom. (0/0)

14
(a/b) She scoured her pots and pans/sink before her company arrived for the lasagne dinner. She always wants to have everything looking spotless. (88.5/0)
(c/d) Anna quickly cleaned the pans/sink before leaving the house for a weekend away. She always wants to have everything looking spotless. (0/0)

15
(a/b) Since the wedding was today, the baker rushed the wedding cake/pies and the bread to the reception. The wedding turned out to be a huge success. (76.9/0)
(c/d) Because it was her sister's big day, Helen made sure the cake/pies and the bread were already at the venue. The wedding turned out to be a huge success. (6.3/0)

16
(a/b) The beautiful model typically shampoos her valuable hair/dogs with a special product. Most of her friends think she exaggerates. (76.9/3.8)

(c/d) The wealthy heiress talks about Jessica's exquisite hair/dogs with a lot of jealousy. Most of her friends think she exaggerates. (0/0)

17
(a/b) The young actress rehearsed her new lines for the play/talk at the university. In the end, she was full of confidence for the event. (65.4/0)
(c/d) The teenager made sure she knew what to do for the play/talk at the community centre. In the end, she was full of confidence for the event. (0/0)

18
(a/b) The famous chef prepared a nice meal/trip for the people who had supported him over the years. Unsurprisingly, they were very grateful to him. (56.7/0)
(c/d) Mr Thompson quickly arranged a meal/trip for the lovely couple that had just become engaged. Unsurprisingly, they were very grateful to him. (0/0)

19
(a/b) The busy bartender gave the regular customer his drink/glass and took his cash. Because he already was well intoxicated, he didn't notice. (46.2/0)
(c/d) One of the temporary staff members gave Marc his drink/glass and took a picture of him. Because he already was well intoxicated, he didn't notice. (0/0)

20
(a/b) When the class went to the zoo, the loud roar of the lion/bear frightened the schoolgirl. She ran away in tears and screaming loudly. (92.0/0)
(c/d) It didn't come as a surprise that the sight of the lion/bear frightened the little girl. She ran away in tears and screaming loudly. (0/0)

21
(a/b) After dinner was completed, the maid washed the dishes/window and then relaxed. The radio was playing some classical tunes. (72.0/0)
(c/d) When he got home, the bachelor stared at the dishes/window and then sighed deeply. The radio was playing some classical tunes. (0/5.6)

22
(a/b) On Sundays, he watches the football game while drinking beer/wine and eating pretzels. He used to be a promising striker, but then he got injured. (85.2/0)
(c/d) Last Friday, Gregory explained to everyone why he liked beer/wine and eating cheese while watching football. He used to be a promising striker, but then he got injured. (0/0)

23
(a/b) The industrious farmer gets all his fresh milk from the cows/goat on his dairy farm. The farm had been in the family for three generations. (70.4/0)
(c/d) My uncle from Texas ultimately decided to get rid of the cows/goat on his ranch. The farm had been in the family for three generations. (0/0)

24
(a/b) She used a needle and thread/string to sew the torn hem in her skirt. She was quite an accomplished seamstress. (81.5/0)
(c/d) Hannah dug up some thread/string to sew the two patches of cloth together. She was quite an accomplished seamstress. (0/0)

25
(a/b) The young nervous paratrooper jumped out of the plane/chair when he heard the shots. While he tried to keep calm, inside he was petrified. (73.1/0)
(c/d) The tired movie maker was sleeping in the plane/chair when he was woken up by a scream. While he tried to keep calm, inside he was petrified. (0/6.7)

26
(a/b) The fat opera singer has meticulously trained her voice/child for a number of years. For some reason, this person found that a real hardship. (59.3/0)
(c/d) The marketeer was told to look after his voice/child for a few hours every day. For some reason, this person found that a real hardship. (6.7/0)

27
(a/b) Her warm sweater was made of wool/silk imported from a village in Northern Ireland. Personally, I think cotton is a much nicer material. (73.3/3.3)
(c/d) At the shop, Lyn asked for some wool/silk imported from New Zealand or Australia. Personally, I think cotton is a much nicer material. (0/0)

28
(a/b) The graceful ballerina pulled some muscles in her legs/arms and could not perform. Apparently she said this in order to get some sympathy. (70.4/0)
(c/d) The woman told us that she disliked her legs/arms and could do with a different pair. Apparently she said this in order to get some sympathy. (0/0)

29
(a/b) The baker put the bread/pizza in the hot oven to bake for 50 minutes. For me, the smell of warm dough is intoxicating. (59.3/0)
(c/d) The woman bought the bread/pizza in the supermarket down the street. For me, the smell of warm dough is intoxicating. (0/0)

30
(a/b) Everyone in the court rose to their feet when the judge/queen entered the room. The room fell silent immediately. (73.1/3.7)
(c/d) We were all taken by surprise when the judge/queen entered the class room. The room fell silent immediately. (0/0)

31
(a/b) At the circus, the audience laughed at the funny clown/tiger running around the stage. Honestly, I found it a sad spectacle. (66.7/0)
(c/d) We were sitting in the front row when we saw the clown/tiger running into a wall. Honestly, I found it a sad spectacle. (0/0)

32
(a/b) The little boy enjoys having some cookies and milk/soda for a snack at night. Obviously, he then had to go brush his teeth. (88.5/0)
(c/d) Apparently, the sweet little boy was given milk/soda for a snack before bed. Obviously, he then had to go brush his teeth. (0/0)

33
(a/b) The minister rang the church bells/chime every Sunday morning before services. The small village was very proud of its medieval church. (63/0)

(c/d) The devout man examined the bells/chime every Saturday after early mass. The small village was very proud of its medieval church. (0/0)

34
(a/b) My younger brother has brilliantly composed a new song/tune for the school play. The composition is very catchy indeed. (63.0/7.4)
(c/d) Last week Monday, Dave's older brother revealed a song/tune for the stag night. The composition is very catchy indeed. (0/0)

35
(a/b) At the wedding, the proud father watched the attractive bride/groom walk down the aisle. The couple had only met ten months before on the internet. (50.0/3.7)
(c/d) Last Saturday, the neighbourhood watched the relaxed bride/groom walk down the road. The couple had only met ten months before on the internet. (0/0)

36
(a/b) The fireman climbed the ladder/stairs to save the desperate people from the flames. The fire had been caused by faulty wiring. (70.4/3.7)
(c/d) Ian carefully took the ladder/stairs to save the child that was overcome by the smoke. The fire had been caused by faulty wiring. (0/0)

37
(a/b) The witness pointed her finger/pencil at the straight-faced defendant. To be honest, I thought that was quite rude. (57.7/0)
(c/d) The caretaker aimed her finger/pencil at the naughty toddler. To be honest, I thought that was quite rude. (0/0)

38
(a/b) The thirsty runner drank a big glass of ice cold water/juice after the marathon. His partner was thirsty as well, but was forgotten. (69.2/0)
(c/d) Without saying anything, Neal went to get some water/juice after the meal. His partner was thirsty as well, but was forgotten. (3.3/0)

39
(a/b) On every job, the photographer takes his camera/sister with him. To be honest, I would have done exactly the same. (50.0/0)
(c/d) I had never expected Jane to take her camera/sister with her. To be honest, I would have done exactly the same. (0/0)

40
(a/b) In the morning, Judy likes to eat oatmeal and poached eggs/meat because of the protein. My vegan friend was horrified when I told her about it. (76.9/3.8)
(c/d) Yesterday evening, my house guest asked for some eggs/meat because she was hungry. My vegan friend was horrified when I told her about it. (0/0)

41
(a/b) The very sleepy baby yawned and rubbed her eyes/nose before falling asleep. It was evening and the rain was tickling against the window. (61.5/7.7)
(c/d) The nurse had to check the patient's eyes/nose before giving him his medicine. It was evening and the rain was tickling against the window. (0/0)

42
(a/b) e saw the old nun and the priest/bishop arrive together for the meeting this morning. Everyone was already seated in the rectory. (56.0/0)
(c/d) To our surprise, we saw the priest/bishop arrive just in time for the annual gathering. Everyone was already seated in the rectory. (0/0)

43
(a/b) When the waiter brought the dinner the banker put salt and pepper/butter on his potato. He likes rich food even though his wife has chastised him numerous times. (92.0/0)
(c/d) After he switched off his computer, the journalist put pepper/butter on his fish pie. He likes rich food even though his wife has chastised him numerous times. (0/0)

44
(a/b) He pounded the nail into the plaster wall with a hammer/wrench and hung the picture. Unfortunately, the wall got damaged in the process. (92.6/0)
(c/d) Last weekend, Peter finally stumbled upon a hammer/wrench and hung the frames on the wall. Unfortunately, the wall got damaged in the process. (0/0)

Appendix B

Experiment 2 items. For each item, the constraining context (a/b/c) appears first. The word before the first / symbol is the predictable target word, the word between the two / symbols is the unpredictable but semantically related word, and the last word in the triplet is the unpredictable and semantically unrelated target word. The same target words appeared in the neutral context (d/e/f). The numbers in parentheses refer to the cloze percentage for the predictable/semantically related/semantically unrelated words, respectively.

1
(a/b/c) For Valentine's day, she received a single red rose/lily/card from the shop around the corner. (96.3/0/0)
(d/e/f) He was careful to choose the right rose/lily/card from the shop around the corner. (0/0/4.0)

2
(a/b/c) The priest wondered how he could get more people to come to the church/sermon/garden even though it was raining. (64.0/4.0/0)
(d/e/f) The widow thought that it was a lovely church/sermon/garden even though it was cold. (0/0/0)

3
(a/b/c) The bomb disposal expert had to remove the right wire/fuse/book without making a mistake. (70.4/0/0)
(d/e/f) Jason was finding it difficult to choose the right wire/fuse/book without making a mistake. (0/0/0)

4
(a/b/c) The young man was deeply saddened when he heard the news/tale/bang at the factory. (88.0/0/0)
(d/e/f) The old man was listening out for the news/tale/bang at the factory. (4.0/0/0)

5
(a/b/c) The girl stuck the two pieces together with glue/tape/rope made by the specialist. (70.4/7.4/0)
(d/e/f) Hal always had plenty of glue/tape/rope made by the specialist. (0/0/0)

6
(a/b/c) In the middle of the cobweb was a black spider/beetle/button that was wet from the rain. (88.9/0/0)
(d/e/f) The girl was surprised to see a large spider/beetle/button that was wet from the rain. (0/0/0)

7
(a/b/c) The comedian's jokes made her laugh/smile/pause for a long time. (64.0/0/0)
(d/e/f) Jane had not been expecting to laugh/smile/pause for a long time. (0/0/0)

8

(a/b/c) The flock of birds flew south for the winter/season/warmth because it was unusually cold. (96.3/0/0)
(d/e/f) The girl was surprised by how much she enjoyed the winter/season/warmth because it was unusually warm. (0/0/0)

9
(a/b/c) The musician opened her door for the man who had come to tune the piano/organ/radio that was in need of repair. (77.8/0/0)
(d/e/f) She was impressed by the craftsmanship that had clearly gone into making the piano/organ/radio that was in need of a home. (0/0/0)

10
(a/b/c) The neighbour asked the dentist to help her with the damage to her tooth/mouth/chair before he went on holiday. (64.0/0/0)
(d/e/f) The boy wished he didn't have to draw a tooth/mouth/chair before he went on holiday. (0/0/0)

11
(a/b/c) Before the battle, the knight tested the weight and balance of his sword/spear/tools that the blacksmith had forged. (66.7/0/0)
(d/e/f) The youth was very pleased with his sword/spear/tools that the blacksmith had forged. (0/0/0)

12
(a/b/c) The young boy baited the mousetraps with cheese/crisps/petals first thing in the morning. (76.0/0/0)
(d/e/f) The baby seemed to be fascinated by cheese/crisps/petals first thing in the morning. (0/0/0)

13
(a/b/c) The cattle would occasionally wander into the nearby field/swamp/hotel behind the church. (68.0/0/0)
(d/e/f) No-one was sure about the size of the field/swamp/hotel behind the church. (0/0/3.7)

14
(a/b/c) The dentist carelessly let the extracted tooth slip from the tweezers into the patient's mouth/beard/shirt after having so much trouble with them. (63.0/0/0)
(d/e/f) John was rather proud of his mouth/beard/shirt after having so much trouble with it. (0/0/0)

15
(a/b/c) Frank was fed up with children throwing stones and breaking his windows/chimney/bicycle as the damage was substantial. (68.0/0/0)
(d/e/f) Richard didn't know how to repair his windows/chimney/bicycle as the damage was substantial. (0/0/0)

16

(a/b/c) Linda was going on holiday, so arranged for a friend to come over and water the plants/garden/rabbit whenever he was able. (76.0/0/0)
(d/e/f) Graham was always calmed by his plants/garden/rabbit whenever he was feeling stressed. (0/0/0)

17
(a/b/c) After dessert, they ordered some coffee/brandy/movies even though it was expensive. (68.0/0/0)
(d/e/f) Joan had heard great things about the coffee/brandy/movies even though it was expensive. (0/0/3.7)

18
(a/b/c) The cat tried its best to catch the mouse/shrew/laser but it was simply too fast. (80.0/0/0)
(d/e/f) The scientist studied the mouse/shrew/laser but it was not yielding interesting results. (0/0/0)

19
(a/b/c) When I went to the park, three men were sitting on the bench/seats/grass next to the wooden gate. (72.0/0/4.0)
(d/e/f) On the outing, Elsie was very impressed by the bench/seats/grass next to the wooden gate. (0/0/0)

20
(a/b/c) Miles won one million pounds by playing the lottery/tombola/trumpet and felt very lucky. (88.9/0/0)
(d/e/f) John was very proud to be put in charge of the lottery/tombola/trumpet and felt very lucky. (0/0/0)

21
(a/b/c) The schoolboy needed to sharpen his pencil/eraser/dagger before the important lesson. (81.5/0/0)
(d/e/f) Mike went to town to buy himself a new pencil/eraser/dagger before the important lesson. (0/0/0)

22
(a/b/c) When I can't sleep, I close my eyes and think of counting sheep/goats/beans until I fall asleep. (84.0/0/0)
(d/e/f) Dale was embarrassed that he'd ordered too many sheep/goats/beans until he realised they'd been running low. (0/0/0)

23
(a/b/c) The builder went to the pub for a pint/beer/chat after a hard day's work. (64.0/0/0)
(d/e/f) Dave was really looking forward to a pint/beer/chat after a hard day's work. (0/7.4/0)

24
(a/b/c) There are twenty-six letters in the alphabet/language/cupboard and they are probably not going to change. (100/0/0)

(d/e/f) For some reason, the child was fascinated by the alphabet/language/cupboard and they are impressed by how quickly she learns. (0/0/0)

25
(a/b/c) Frank always makes sure to vote in order to support his political party/group/cause at the elections. (74.1/0/0)
(d/e/f) Jo was convinced that it was a very good party/group/cause at the elections. (0/0/0)

26
(a/b/c) The teenager put on his headphones and listened to his music/songs/mates for his whole journey. (88.0/0/0)
(d/e/f) Alan would not have been able to get through the week without his music/songs/mates for his own entertainment. (4.0/0/0)

27
(a/b/c) The greedy businessman was obsessed with money/coins/rugby and always had been. (72.0/0/0)
(d/e/f) Mary was far too interested in money/coins/rugby and always had been. (0/0/0)

28
(a/b/c) To succeed, it was important for everybody to work as a team/unit/cook for the rest of the week. (92.6/0/0)
(d/e/f) Mike was sure that they were going to have to get a new team/unit/cook for the rest of the week. (0/0/0)

29
(a/b/c) The shepherd spent all day looking for his lost sheep/lambs/phone he had only acquired last week. (92.0/0/4.0)
(d/e/f) Alan was missing his sheep/lambs/phone he had only acquired last week. (0/0/4.0)

30
(a/b/c) Tom was unable to go on; he felt drained and had completely run out of energy/breath/rivets but the work was not yet finished. (60.0/0/0)
(d/e/f) George was surprised to find he was out of energy/breath/rivets but the work was not yet finished. (0/0/0)

31
(a/b/c) To complete his Halloween costume, the young boy wore a scary mask/cape/grin and frightening make-up. (84.0/0/0)
(d/e/f) Lucy hadn't expected to see her brother with a mask/cape/grin and frightening make-up. (0/0/0)

32
(a/b/c) Princess Anne felt confident as she mounted the horse/camel/steps that had been provided. (81.5/0/3.7)
(d/e/f) James was not satisfied with the height of the horse/camel/steps that had been provided. (0/0/0)

33
(a/b/c) He lit the candle with a match/taper/flame to light the way. (84.0/0/0)
(d/e/f) The girl took great care whilst holding the match/taper/flame to light the way. (0/0/0)

34
(a/b/c) The chef couldn't help crying as he chopped up the onions/carrot/planks with a lot of anger. (80.0/8.0/0)
(d/e/f) Simon urgently had to prepare the onions/carrot/planks with a lot of haste. (0/0/0)

35
(a/b/c) The cut on the boy's head was very large, so he was going to need stitches/bandages/medicine as soon as possible. (80.0/0/0)
(d/e/f) Fiona was asked to order some more stitches/bandages/medicine as soon as possible. (0/0/0)

36
(a/b/c) The lovers lay in the field at night, gazing up at the stars/skies/plane until it got too cold. (88.0/0/0)
(d/e/f) Sally loved to watch the stars/skies/plane until it got too cold. (4.0/0/0)

37
(a/b/c) The kids were going to the pond, and took some bread for the ducks/birds/lunch at the weekend. (100/0/0)
(d/e/f) The boy was particularly looking forward to the ducks/birds/lunch at the weekend. (0/0/0)

38
(a/b/c) The cook stirred the stew with a large spoon/whisk/broom with a long handle. (60.0/0/0)
(d/e/f) Malcolm promised to replace the spoon/whisk/broom with a long handle. (0/0/0)

39
(a/b/c) The doctor told Fred that his drinking would damage his liver/heart/skill unless he was more careful. (59.3/0/0)
(d/e/f) Yesterday Fred told his friend that they will look at his liver/heart/skill unless he was more careful. (0/0/0)

40
(a/b/c) The banker loaned the businessman some more money/shares/tools to build a new shop. (81.5/0/0)
(d/e/f) The neighbour across the road wanted to get money/shares/tools to build a new shop. (0/0/0)

41
(a/b/c) For Halloween, Liz dressed up as an ugly old witch/ghost/nurse and was very excited. (55.6/0/0)
(d/e/f) For her arts class, the little girl drew a witch/ghost/nurse and was very excited. (0/0/0)

42

(a/b/c) The baby laughs and giggles when she shakes her new rattle/bottle/gloves and everybody claps for her. (80.0/0/0)
(d/e/f) Last weekend, good friends of ours bought us a new rattle/bottle/gloves and everybody has noticed it. (0/8.0/0)

43
(a/b/c) The skilled gardener went outside to pull up the weeds/roses/fence before it started raining. (88.0/0/0)
(d/e/f) The distracted professor even didn't notice the weeds/roses/fence before it started raining. (0/0/0)

44
(a/b/c) He is a grouch in the morning until he has had his coffee/cereal/shower of some description. (56.0/0/0)
(d/e/f) Carl told me that he wanted some time to have a coffee/cereal/shower of some description. (0/0/7.4)

45
(a/b/c) The maid was told to wash and wax the kitchen floor/tiles/table quickly and quietly. (66.7/0/7.4)
(d/e/f) The tired young woman was informed that the floor/tiles/table quickly needed cleaning. (0/0/0)

46
(a/b/c) The famous chef prepared a nice meal/dish/trip for the happy couple. (72.0/4.0/0)
(d/e/f) Mr Thompson quickly arranged a meal/dish/trip for the happy couple. (3.7/0/0)

47
(a/b/c) When the class went to the zoo, the loud roar of the lion/bear/wind in the grass scared the children. (92.0/0/0)
(d/e/f) It didn't come as a surprise that the sight of the lion/bear/wind in the grass scared the children. (0/0/0)

48
(a/b/c) After dinner was completed, the maid washed the dishes/plates/window until the telephone rang. (76.0/8.0/0)
(d/e/f) When he got home, the bachelor stared at the dishes/plates/window until the telephone rang. (0/0/3.7)

49
(a/b/c) The industrious farmer gets all his fresh milk from the cows/goat/shop that he owns. (72.0/0/8.0)
(d/e/f) My uncle from Texas ultimately decided to get rid of the cows/goat/shop that he owned. (4.0/0/0)

50
(a/b/c) She used a needle and thread/string/shears to fix the dress. (84.0/0/0)

(d/e/f) Hannah dug up some thread/string/shears to fix the dress. (0/0/0)

51
(a/b/c) The young nervous paratrooper jumped out of the plane/train/chair when he heard the announcement. (66.7/0/0)
(d/e/f) The tired movie maker was sleeping in the plane/train/chair when he heard the announcement. (0/0/8.0)

52
(a/b/c) The fat opera singer has meticulously trained her voice/sound/child for a number of years. (68.0/0/0)
(d/e/f) The marketer was told to look after his voice/sound/child for a number of years. (0/0/4.0)

53
(a/b/c) The graceful ballerina pulled some muscles in her legs/arms/show quite badly and was in pain. (59.3/3.7/0)
(d/e/f) The woman told us that she disliked her legs/arms/show quite a lot and was feeling down. (0/0/0)

54
(a/b/c) The baker put the bread/cakes/shelf in the shop window. (80.0/4.0/0)
(d/e/f) The woman bought the bread/cakes/shelf in the shop window. (0/0/0)

55
(a/b/c) The little boy enjoys having some cookies and milk/coke/hugs as a special treat. (66.7/0/0)
(d/e/f) Apparently, the sweet little boy was given milk/coke/hugs as a special treat. (0/0/0)

56
(a/b/c) The minister rang the church bells/chime/phone in a real hurry. (76.0/0/4.0)
(d/e/f) The devout man examined the bells/chime/phone in a real hurry. (0/0/0)

57
(a/b/c) The fireman climbed the ladder/stairs/cliffs all the way to the top. (56.0/0/0)
(d/e/f) Ian carefully took the ladder/stairs/cliffs all the way to the top. (0/0/0)

58
(a/b/c) The very sleepy baby yawned and rubbed her eyes/nose/doll very carefully. (80.0/4.0/0)
(d/e/f) The nurse had to check the patient's eyes/nose/doll very carefully to keep the child happy. (0/0/0)

59
(a/b/c) To complete the meal, the chef put salt and pepper/nutmeg/spoons on the table. (92.6/0/0)
(d/e/f) After he switched off his computer, the journalist put pepper/nutmeg/spoons on the table. (0/0/0)

60

(a/b/c) He pounded the nail into the plaster wall with a hammer/wrench/bottle and hurt himself in the process. (96.0/0/0)
(d/e/f) Last weekend, Peter finally stumbled upon a hammer/wrench/bottle and hurt himself in the process. (0/0/0)