

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/318649626>

# How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study.

Conference Paper · July 2017

CITATIONS

0

READS

248

2 authors, including:



Aine Ito

University of Oxford

15 PUBLICATIONS 34 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Prediction in language comprehension [View project](#)

# How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study.

Aine Ito †, and E. Matthew Husband ‡

Faculty of Linguistics, Philology & Phonetics, University of Oxford  
Clarendon Institute, Walton Street, Oxford, OX1 2HG, United Kingdom

E-mail: † aine.ito@ling-phil.ox.ac.uk, ‡ matthew.husband@ling-phil.ox.ac.uk

**Abstract** Prediction is often assumed to play a crucial role during language comprehension. While some theories propose that prediction robustly affects at all levels of linguistic representation, empirical evidence suggests that the circumstances under which linguistic predictions occur appear to be limited, particularly when comparing prediction of phonological information to semantic information. To more directly explore these limits, we compared effects of semantic and phonological prediction in a visual world eye-tracking study. Participants heard sentences where the target word was either predictable (e.g., “*That dog looks so happy, wagging its tail ...*”) or unpredictable (“*If there is one, click on the picture of the tail.*”), while viewing objects that corresponded to either the target word (*tail*), a semantic competitor word (*paw*), a phonological competitor word (*table*), or an unrelated word (*daisy*). Target and semantic competitor objects attracted more fixations than unrelated objects well before the target word onset in predictable sentences and not in unpredictable sentences, suggesting that participants predicted semantic information. However, there were no predictive eye movements for phonological competitor objects. The results suggest that phonological prediction is not as robust as semantic prediction.

**Key words** Prediction, Comprehension, Eye-tracking, Visual world paradigm

## 1. Introduction

Many current theories of language processing have argued that language comprehension involves predictive processes (e.g., Dell & Chang, 2014; Pickering & Garrod, 2013). Visual world eye-tracking studies have shown that people predict some information about upcoming words (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Rommers, Meyer, Praamstra, & Huettig, 2013). Evidence for prediction has been found in very young children (Mani & Huettig, 2012), and sometimes in non-native speakers (Ito, Corley, & Pickering, 2017). Event-Related Potential (ERP) studies also find evidence for prediction using different designs (Federmeier & Kutas, 1999; Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort, 2005).

However, predictions do not occur consistently at all levels. Available evidence for semantic prediction is well attested, but evidence for phonological prediction is scarce and mixed. In perhaps the most well known case for phonological prediction,

DeLong, Urbach, and Kutas (2005) utilised the phonological rule of English *a/an* articles, and had participants read these articles that either matched or mismatched an expected noun (e.g., “*The day was breezy so the boy went outside to fly a kite/ an airplane...*”). They found a correlation between the article N400s and the articles’ offline predictability (measured in a cloze test), and took this finding to argue that people probabilistically pre-activate phonological information of upcoming words. However, a recent large-scale study involving nine labs failed to replicate this graded effect for articles (Nieuwland et al., 2017). Additionally, Martin et al. (2013) used the same logic as in DeLong et al. (2005), and found a larger negativity in a typical N400 window for unexpected articles compared to expected articles. However, this difference between the conditions was not found in the original DeLong et al.’s study (they only found a correlation between N400 and cloze probability for articles). Moreover, neither this condition effect nor a graded effect of article cloze was conceptually replicated in two experiments in Ito, Martin, and Nieuwland (2016), much like

Nieuwland et al (2017). Thus, evidence for phonological prediction from ERP studies using this *a/an* manipulation is actually far from clear-cut.

In an ERP study, Ito, Corley, Pickering, Martin, and Nieuwland (2016) investigated prediction of semantic and phonological/orthographic form information. Participants read sentences containing a predictable word at a standard 500ms or slower 700ms Stimulus Onset Asynchrony (SOA). At both SOAs, words that were semantically related to the predictable word elicited a smaller N400 effect compared to words that were unrelated to predictable words, suggesting that participants predicted semantic information of the predictable word. In contrast, words that were form-related to the predictable word showed this N400 reduction only at the 700ms SOA, and only in very high-cloze sentences (Mean cloze = 94%). These findings suggest that prediction of a specific word form is less likely to occur compared to prediction of semantic information, though it is possible that semantically related words in this study were simply more strongly related to the predictable word than form-related words.

In addition to evidence from ERPs, one visual world eye-tracking study utilising a phonological competitor effect found evidence for phonological prediction (Ito, Pickering, & Corley, submitted). In this study, participants listened to sentences containing a highly predictable word, and viewed a scene that contained an object representing the predictable word (e.g., *cloud*), a phonological competitor word (e.g., *clown*), or an unrelated word (e.g., *globe*). Before hearing the predictable word, participants were more likely to fixate the phonological competitor object as well as the predictable object relative to the unrelated object. The findings demonstrate that participants predicted phonological information of predictable words. However, in an aim to maximise the likelihood of detecting a phonological prediction effect, this study used slowly spoken sentences with near-100% cloze probabilities. Despite this prediction-optimised design, the phonological prediction effect itself was short-lived compared to more long-lasting effects of semantic prediction (prediction of shape information) in a study using a very similar design (e.g., Rommers et al., 2013). Thus, it is possible that phonological prediction is very weak, and occurs only

under limited or highly atypical circumstances.

Overall, evidence for prediction of phonological information has been quite unreliable compared to semantic information, often requiring a stimulus that is much slower than typical speech. This suggests a dilated time course for the prediction of phonological information. However, it is unclear whether phonological prediction is strictly weaker or merely delayed compared to semantic prediction. To address this issue and add further evidence to the debate, we followed a competitor design (e.g. Ito et al., submitted), and compared semantic and phonological competitor effects in predictable sentences and in unpredictable sentences. Since prediction-driven eye movements are thought to be closely time-locked to the point when predictions occur, the visual world investigation allowed us to test whether the semantic and phonological predictions simply differ in their time-course.

## 2. Methods

### 2.1. Participants

Participants were 23 native British English speakers studying at the University of Oxford (18 females, age  $M = 23$  years, range = 19-30 years). All participants had normal or corrected vision.

### 2.2. Stimuli

Our auditory stimuli were 56 predictable-unpredictable sentence pairs. Predictable sentences contained a highly predictable target word (e.g., *tail* in “*That dog looks so happy, wagging its tail as it walks along.*”). Unpredictable sentences contained the same target word in the context “*If there is one, click on the picture of the ...*”, so that (un)predictability was constant for all items. We tested the predictability of the target words in predictable sentences in a cloze probability test. Forty native British English speakers read predictable sentence contexts truncated before the target word (e.g., “*That dog looks so happy, wagging its*”) and completed each of them with the first word that came to mind. The mean cloze probability of the target word was 95% ( $SD = 6$ , range = 80-100%). We created additional 56 filler sentences, half with various contexts, and the rest with the unpredictable sentence context described above. We created fillers this way because we divided the

experiment into predictable and unpredictable blocks, and we wanted to ensure that each block contained the same number of fillers (50% of the trials) with similar or same sentence structures. These sentences were recorded by a female native British English speaker, who read the sentences clearly and slowly (511 ms per word on average).

Our visual stimuli consisted of four (one critical and three distractor) objects. The critical object represented the target word (e.g., *tail*) in the target condition, a semantic competitor word (e.g., *paw*) in the semantic condition, a phonological competitor word (e.g., *table*) in the phonological condition, and an unrelated word (e.g., *daisy*) in the unrelated condition. The distractor objects were the same across conditions. Because objects that are more plausible to be mentioned are likely to be fixated over those that are less plausible and fixation bias towards a particular object could, at least partially, reflect people’s preference to look at a plausible object, we created the sets so that all objects except the target object were implausible in the predictable sentence context. Therefore, any fixation bias towards semantic and phonological competitors was more likely to be driven by information overlap between predictable target words and these competitor words. Semantic competitor words had varied relationships with the target word; some belonged to the same semantic category (e.g., *nose-ear*), some had a whole-part relationship (e.g., *car-tyre*), and others frequently co-occur (e.g., *tea-lemon*). Participants for the eye-tracking study assessed semantic relatedness between target objects and non-target critical objects (semantic competitor, phonological competitor and unrelated objects) after the eye-tracking experiment. On the 7-point scale (1 = not related at all, 7 = highly related), the mean semantic relatedness rating was 5.2 ( $SD = 1.6$ ) for the semantic condition, 1.9 ( $SD = 1.3$ ) for the phonological condition, and 1.6 ( $SD = 1.1$ ) for the unrelated condition. Phonological competitor words shared phonological onset with the target word (Mean phoneme overlap = 2.3). Majority of them were also orthographically related to the target word. The object pictures were mostly taken from the Bank of Standardized Stimuli (Brodeur, Guérard, & Bouras, 2014), and some from freely available ClipArt.

Eight experimental lists were created by crossing condition (target, semantic, phonological, and unrelated) and block order (predictable and unpredictable). Participants got only one variant of each item. Critical objects appeared at each of the four quadrants equally frequently. An example of the visual stimuli for each condition is shown in Figure 1.



**Figure 1.** An example item for the predictable sentence “*That dog looks so happy, wagging its tail as it walks along.*” and for the unpredictable sentence “*If there is one, click on the picture of the tail.*”

### 2.3. Procedure

Participants’ eye movements were recorded using an EyeLink 1000 Desktop mount eye-tracker sampling at 500 Hz. Participants were instructed to listen to the sentences, and click on an object that was mentioned in the sentence, or click the background if none of the objects were mentioned. In each trial, the visual scene first appeared on the screen, and 3000ms later the sentence was presented via headphones. The pictures were presented on a viewing monitor at a resolution of 1024×768 pixels.

After the eye-tracking experiment, participants performed a semantic rating task described in the Stimuli section. Participants saw pairs of objects and rated how strongly the two pictures were related in meaning. The experiment took about 50 minutes in total.

## 3. Results

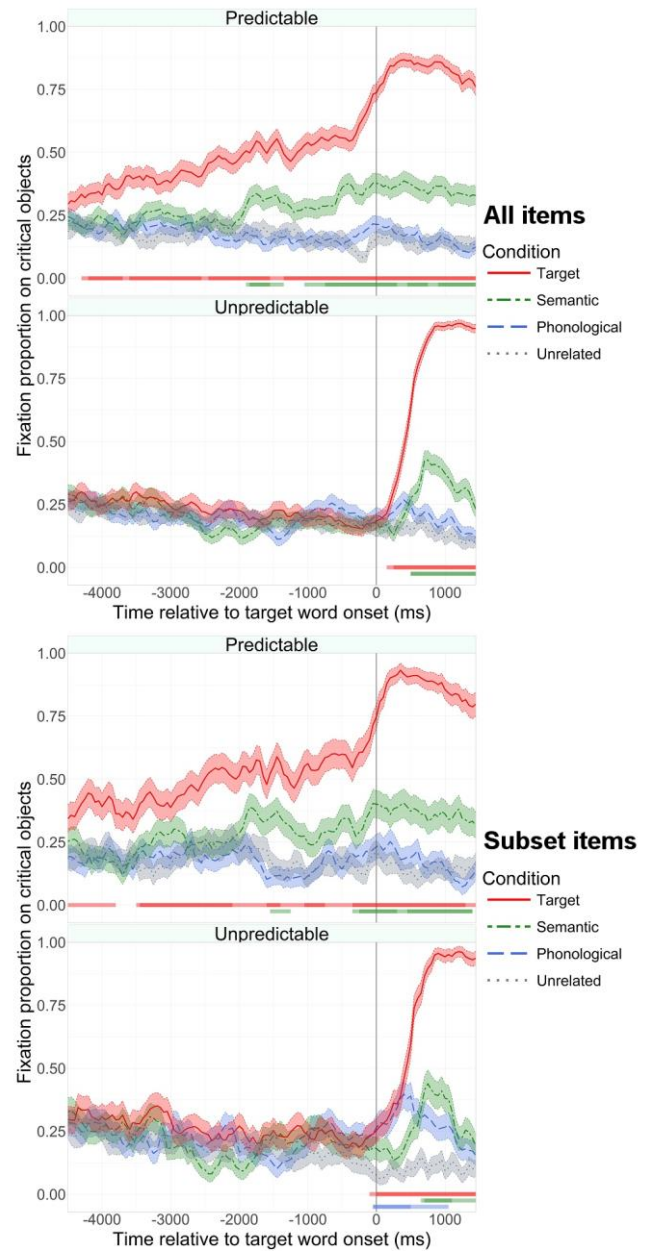
### 3.1. Eye-tracking analysis

We first calculated the proportion of time spent fixating on critical objects for each 50ms time bin relative to the target word onset. Bins containing blinks were excluded. We transformed fixation proportion to log odds using the empirical logit function (Barr, 2008), and used this as a dependent variable. We analysed fixation data from -4500ms to 1500ms relative to the target word onset using a cluster randomisation technique (Kronmüller, Noveck,

Rivera, Jaume-Guazzini, & Barr, 2017). This technique can identify time ranges in which an effect is statistically significant while correcting for multiple comparisons over time bins and avoiding the arbitrariness in choosing the size of analysis time bins. The basic logic is as follows: (1) We first identified “clusters” by grouping subsequent time bins in which all effects are significant with  $p < .05$  and are in the same direction. (2) For each of these clusters, we calculated a “cluster mass statistic” – the sum of all of the individual test statistics in the cluster. (3) We then permuted the data by shuffling the condition labels. Because the condition was a within-subject variable, we shuffled it within-subjects. (4) We created 9999 alternative versions of the dataset using this permutation scheme, and performed (1) and (2) on each dataset. The distribution of cluster mass statistics over the transformed datasets provide a null-hypothesis distribution, against which the cluster-mass statistics from the original dataset was compared. The  $p$ -values were determined by comparing the log odds for the original dataset to the distribution comprised by all other datasets. The analysis compared target vs. unrelated conditions, semantic vs. unrelated conditions, and phonological vs. unrelated conditions in predictable and unpredictable sentences separately. We performed by-subject and by-item analyses (cf. Kronmüller et al., 2017).

### 3.2. Eye-tracking results

The results for all items are shown in Figure 2 (top). Target objects started to attract more fixations than unrelated objects from around 4500ms before the target word onset in predictable sentences, and only after the target word onset in unpredictable sentences. Semantic competitor objects started to attract more fixations than unrelated objects from around 2000ms before the target word onset in predictable sentences, and from around 500ms after the target word onset in unpredictable sentences. The fixation proportion on phonological competitor and unrelated objects did not differ in any of the time windows. Thus, we found evidence for semantic prediction, but we did not find any evidence for phonological prediction.



**Figure 2.** Results for all items (top panel) and for 28 subset items (bottom panel). Fixation proportion on critical objects in the predictable and unpredictable sentences. The lines at the bottom of each plot indicate significant differences between target and unrelated (red), semantic and unrelated (green), and phonological and unrelated (blue) conditions, corresponding to the time range on x-axis. Lines with a darker shade indicate that the difference was significant in by-subject and by-item analyses, and lines with a lighter shade indicate that the difference was significant only in by-subject or by-item analysis.

However, though not significant, there is a small fixation proportion difference between the phonological and unrelated conditions around -500ms to 0ms relative to the target word onset in predictable sentences. Moreover, since we did not find a phonological competitor effect in unpredictable sentences, our items might not have a power to detect a phonological prediction effect. To explore this possibility, we performed the same analyses on a subset of 28 items in which the phonological competitor attracted more fixations than the unrelated object in unpredictable sentences between 0-1000ms. Fixation proportion in these items is presented in Figure 2 (bottom). In this subset, there was a clear phonological competitor effect in unpredictable sentences, which peaked slightly earlier than the semantic competitor effect, and was similarly robust to the semantic competitor effect. Critically, however, there was no phonological competitor effect in predictable sentences. Thus, we found no evidence for phonological prediction as measured by phonological competition in the visual world paradigm.

#### 4. Discussion

We found predictive eye movements to predictable target objects and semantic competitor objects, but we found no phonological competitor effect in predictable sentences. These results suggest that phonological prediction is not as robust as semantic prediction, even at relatively high and potentially atypical cloze probabilities.

Our findings largely fit with the available evidence. In most of the visual world studies that found predictive looks to upcoming target objects (e.g., Altmann & Kamide, 1999), it is possible that people did not predict phonological information, or phonological information was predicted only when the predictable target was visually present. Since phonological prediction likely involves prediction of a specific word, our results suggest that prediction of a specific word can fail to occur even when the word is highly predictable. In natural language, predictable words are very rare. An estimate in Luke and Christianson (2016) suggests that only about 5% of content words have a cloze probability that is higher than 67%. Our sentences had a mean cloze probability of 95%, and such highly predictable sentences would occur even less frequently.

Thus, people may predict broad semantic information (e.g., something related to a dog), but may be hard pressed to predict more detailed phonological information of a specific word.

Why did we not find any evidence for phonological prediction in contrast to Ito et al. (submitted)? One explanation may be a difference in preview time. The objects were presented 1000 ms before predictable words in Ito et al., in contrast to 3000 ms before the sentence onset in the current study. Although this gave people more time to encode the visual objects (Ferreira, Foucart, & Engelhardt, 2013), a greater competitor effect may emerge when the picture onset is closer to the point of the target word that they are predicting. Alternatively, our sentences had slightly lower cloze probabilities (95% in all items, 94% in the subset) than those in Ito et al. (98%). It is possible that this small difference eliminated the effect, but if so, it fits with our conclusion, namely, phonological prediction is not particularly robust.

One might argue that the phonological competitor effect in unpredictable sentences in the subset data may not be real. We think this is unlikely, because the time-course of phonological and semantic competitor effects was similar to a previous study (Huettig & McQueen, 2007), which found an earlier phonological competitor effect compared to semantic competitor effect in neutral sentence context (Experiment 1).

In conclusion, our findings support the hypothesis that phonological prediction is very limited compared to semantic prediction. We conjecture that conditions that allow phonological prediction to occur may rarely be available during natural language comprehension, and propose that the role of prediction, particularly that of specific word, must be carefully evaluated.

#### 5. Acknowledgements

We thank Hannah Davidson and Sherry Yong Chen for research assistance.

#### 6. References

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

- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase ii: 930 new normative photos. *PLoS ONE*, 9(9).
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B*, 369, 20120394.
- 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.
- 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, 469–495.
- Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, 69(3), 165–182.
- Huetting, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57(4), 460–482.
- Ito, A., Corley, M., & Pickering, M. J. (2017). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, 1–14.
- Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157–171.
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*.
- Ito, A., Pickering, M. J., & Corley, M. (submitted). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.
- Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., & Barr, D. (2017). The positive side of a negative reference: the delay between linguistic processing and common ground. *Royal Society Open Science*, 4(2).
- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843–847.
- Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4), 574–588.
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... Huettig, F. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *bioRxiv*, 111807.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioural and Brain Sciences*, 36, 329–392.
- Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51(3), 437–447.
- 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.