

Chow, W.Y., & Chen, D. (2020). Predicting (in)correctly: listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience*. doi: 10.1080/23273798.2020.1733627

**Predicting (in)correctly: listeners rapidly use unexpected  
information to revise their predictions**

Wing-Yee Chow & Di Chen

*Division of Psychology and Language Sciences, University College London, UK*

Corresponding author:

Wing-Yee Chow (wingyee.chow@ucl.ac.uk)

Chandler House, 2 Wakefield Street, London, WC1N 1PF, United Kingdom

## **Abstract**

Comprehenders can incorporate rich contextual information to predict upcoming input on the fly, and cues that conflict with their predictions are quickly detected. The present study examined whether and how comprehenders may revise their existing predictions upon encountering a prediction-inconsistent cue. We took advantage of the rich classifier system in Mandarin Chinese and tracked participants' eye-movements as they listened to sentences in which the final noun is preceded by a classifier which was either compatible with the most expected noun, incompatible with the most expected noun but indicative of another contextually suitable noun, or uninformative. We found that, upon hearing a prediction-inconsistent classifier, listeners quickly directed their eye gaze away from the originally expected object and immediately onto the (initially) unexpected but contextually suitable object. This provides initial evidence that listeners can quickly use prediction-mismatching cues to revise their existing predictions on the fly.

**Keywords:** real-time language comprehension; prediction; prediction error; prediction revision; eye-tracking

## **Introduction**

The human brain has an impressive capacity to process information in real time, and many have attributed this to its ability to anticipate upcoming events on the fly. Over the past two decades, predictive processes have been demonstrated across different domains of cognition, such as visual and auditory perception (e.g., Summerfield & Egner, 2009; Bendixen, Schroger, & Winkler, 2009), motor planning (e.g., Wolpert & Flanagan, 2001), and language comprehension (for a review see Kamide, 2008 and Kutas, DeLong, & Smith, 2011). The present study is concerned with the cognitive processes that take place when we encounter cues that are inconsistent with our existing prediction.

In the domain of language, many studies have shown that comprehenders can incorporate rich contextual information to generate predictions on the fly. The facilitative effects of prediction on real-time language comprehension have been demonstrated in studies using different experimental techniques such as eye-tracking and event-related brain potentials (ERPs). For example, in a classic “visual world” eye-tracking study, Altmann and Kamide (1999) presented participants with a visual display with multiple objects and recorded their eye-movements on the visual display whilst they listened to utterances. They found that, when presented with an utterance like “The boy will eat the cake” along with a visual display depicting a cake among non-edible objects, listeners were more likely to look at the cake (the only edible object on the display) even before the word “cake” was uttered compared to a neutral utterance like “The boy will move the cake,” showing that comprehenders can use contextual information such as a verb’s selectional restriction to quickly anticipate upcoming referents. Meanwhile, studies measuring comprehenders’ ERPs have shown that the size of the N400 response, a negative-going response that peaks at around 400 ms post-stimulus onset, is reduced when a word is more predictable (e.g., Kutas & Hillyard, 1984). Based on

the view that the N400 reflects the process of accessing long-term semantic memory representations (e.g., Federmeier & Kutas, 1999; Lau, Poeppel & Phillips, 2008), the N400's sensitivity to a word's predictability has been taken to show that access to long-term semantic memory is facilitated when the semantic representations associated with a word may be pre-activated as a result of prediction (e.g., Chow, Lau, Wang, & Phillips, 2018; Chow, Smith, Lau, & Phillips, 2016; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Szewczyk & Schriefers, 2018; Wlotko & Federmeier, 2012).

### ***When prediction goes wrong***

However, the expressive nature of language means that our predictions often turn out to be wrong, and comprehenders may encounter evidence that their predictions are incorrect in at least two different ways.

In some cases, comprehenders can detect a prediction error when the input outright violates their predictions. For example, comprehenders may predict a noun like *'cake'* upon reading "On my birthday my grandma always bakes me a ...," and this prediction would be fully violated if another noun like *'loaf'* appeared instead. In situations like this, the unexpected word directly negates comprehenders' predictions and forces them to change their interpretation of the sentence. Such outright prediction violations have long been documented to have a disruptive effect on processing. For example, a number of ERP studies have shown that unexpected but congruous words elicited a larger late frontal positivity when the preceding context is strongly predictive of another word (e.g., DeLong, Quante, & Kutas, 2014; Federmeier et al., 2007), and this late frontal positivity has been proposed to index the inhibition of the original prediction (Kutas, 1993; for a review see Van Petten & Luka, 2012).

In other cases, however, the input does not directly negate an existing prediction, but it may simply signal to comprehenders in advance that their existing prediction is likely to be wrong. For example, upon reading “This book has three ...,” one might expect to see a noun like “volumes” or “editions” next; however, if the sentence continues with a numeral like “hundred,” then it might indicate to the reader that his or her original predictions are likely to be wrong, and nouns like “pages” might become more likely instead. Crucially, unlike cases in which an existing prediction is outright violated by an unexpected word (see above), here the input provides an early sign that an existing prediction may be wrong. We will focus on cases like this in the present study.

Previous studies across different languages have shown that comprehenders can reliably detect early signs that their prediction may be wrong. For example, Van Berkum and colleagues (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) measured participants’ ERPs as they listened to sentences in Dutch such as ‘The burglar had no trouble locating the secret family safe. Of course, it was situated behind a ...’ in which a specific noun (e.g., *painting*<sub>NEU</sub>) is highly predictable. Crucially, when the sentences were continued with an adjective with an inflectional suffix that marked a different grammatical gender than that of the predicted noun (e.g., ‘grote’ = *big*<sub>COM</sub> instead of ‘groot’ = *big*<sub>NEU</sub>), an early positivity was observed at the adjective, suggesting that comprehenders can very quickly detect a mismatch between the grammatical gender information carried by the inflectional suffix *-e* on the adjective and the noun they are expecting.

This experimental paradigm (which we will call a ‘pre-target mismatch paradigm’) has been used to examine comprehenders’ sensitivity to various pre-target cues that can signal potential prediction errors in different languages. To date, this paradigm has been used to study noun prediction with gender-marked determiners in Spanish (Wicha, Bates, Moreno &

Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004), gender-marked determiners and adjectives in Dutch (Kochari & Flecken, 2018; Otten, Nieuwland & Van Berkum, 2007; Otten & Van Berkum, 2008, 2009; Van Berkum et al., 2005), animacy-marked determiners and adjectives in Polish (Szewczyk & Schriefers, 2013; Szewczyk, 2018), indefinite articles and adjectives in English (Boudewyn, Long & Swaab, 2015; DeLong et al, 2005; cf. Ito, Martin & Nieuwland, 2017; Nieuwland et al., 2018), and nominal classifiers in Mandarin Chinese (Kwon, Sturt & Liu, 2017). Across these studies, pre-nominal cues that were inconsistent with an expected noun commonly elicited differential ERP responses even before the identity of the noun itself was revealed.<sup>1,2</sup> Taken together, these findings suggest that comprehenders can anticipate upcoming words on the fly and detect early signs that are inconsistent with their predictions.

### ***Revising predictions as a race against time***

These data, however, do not tell us *what cognitive processes are engaged* when we encounter cues that are inconsistent with our predictions. That is, even though previous studies have commonly observed ERP effects at prediction-inconsistent pre-nominal cues, little is known about what cognitive processes take place when prediction-consistent cues are detected – do

---

<sup>1</sup> This is true for all cases in which the pre-nominal cue depends on the semantic and/or syntactic features of the upcoming noun. However, in the case of indefinite articles ‘a/an’ in English which are dependent on the phonological property of the following word (which needs not be a noun, e.g., ‘an orange kite’), the reliability of the effect (first reported by DeLong et al., 2005) has been contested after multiple failed replication attempts (Ito et al. 2017; Nieuwland et al., 2018). We will not expand on this further since the classifier-noun dependency in Mandarin Chinese is not phonological in nature.

<sup>2</sup> The ERP effects reported by these studies varied in their polarity, latency, and scalp distribution (see Kochari & Flecken, 2018 for a summary). However, a discussion about such variations is beyond the scope of the present paper.

comprehenders actively inhibit their original predictions and make new predictions? Or are there circumstances under which they simply wait until the relevant bottom-up inputs to arise?

On the one hand, the received wisdom that language processing is highly incremental strongly favours the view that disconfirmed predictions are revised immediately. Previous studies have demonstrated that comprehenders can quickly integrate various sources of information and constraints to process language incrementally (e.g., Kamide, Altmann, & Haywood, 2003; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Tanenhaus, & Garnsey, 1994; c.f. Clifton et al., 2003). Further, research on the processing of temporary syntactic ambiguities in “garden-path” sentences has shown that syntactic reanalysis is triggered as soon as comprehenders encounter inputs that are inconsistent with their existing analysis (Frazier & Rayner, 1982). Therefore, under this view, comprehenders are expected to revise their predictions immediately upon encountering disconfirming evidence in the input.

On the other hand, while reanalysis is necessary for comprehenders to successfully interpret garden-path sentences, prediction revision is not necessary for successful comprehension because the relevant input will become available to the comprehenders eventually. That is, even if comprehenders do not revise their predictions upon encountering prediction-inconsistent cues, they will still be able to process the relevant input when it arises. In addition, information that disconfirms a strong prediction may have a disruptive effect on processing (e.g., Federmeier et al., 2007; van Petten & Luka, 2012; c.f. Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016) and may prevent comprehenders from immediately revising their predictions. Further, it may not be worthwhile for the language processing system to revise disconfirmed predictions in some cases. This is because the process of revising predictions presumably takes time and cognitive resources (including cognitive

control in inhibiting an existing prediction and working memory in computing new predictions), and yet the relevant input may become available immediately after the prediction-inconsistent cue, possibly even before a revised prediction can be computed. Under this view, we may not find evidence for prediction revision even when comprehenders can detect cues that are inconsistent with their predictions in the input.

To date, studies that used the pre-target mismatch paradigm have commonly observed an N400 effect at the noun, suggesting that the unexpected noun was more difficult to process than the expected noun even when it was preceded by a prediction-inconsistent cue. However, it has remained unclear whether the prediction-inconsistent cue makes the unexpected noun itself any more expected. For example, Van Berkum et al.'s (2005) observation of a larger N400 response to the unexpected noun ('bookcase') than to the expected noun ('picture') suggested that 'bookcase' was still more difficult to process than 'picture' even when its gender had already been signalled by a preceding adjective, but it could not tell us whether the unexpected noun 'bookcase' became any less unexpected due to the presence of a gender-marked adjective. In order to ask this question, a comparison must be made against a baseline condition in which the cue preceding the unexpected noun is not prediction-inconsistent. We believe that the rich nominal classifier system in Mandarin Chinese provides an ideal testing ground for asking this question. Below we will first review some key properties of Mandarin Chinese nominal classifiers which make them suitable for our investigation into prediction revision.

### ***Nominal Classifiers in Mandarin Chinese and Prediction***

In Mandarin Chinese, when a noun is modified by a numeral (e.g., one, two) or a demonstrative (e.g., this, that), it must be preceded by a grammatical element known as a "classifier" (Li & Thompson, 1989).

(1) 三 杯 咖啡

three CL<sub>bēi</sub> coffee

‘three cups of coffee’

(2) 一 朵 花

one CL<sub>duǒ</sub> flower

‘a flower’

Mandarin Chinese has over 70 nominal classifiers (Erbaugh, 2004), and a majority of them can only be used with nouns that share certain semantic-conceptual properties like shape and function (e.g., ‘zhī’ (枝) for pens and tree branches, ‘dǐng’ (顶) for hats and crowns; Allan, 1977; Tai, 1994). These are known as specific classifiers because they are associated with specific classes of nouns. Meanwhile, a small number of classifiers such as ‘gè’ (个) and ‘xiē’ (些) are compatible with a wide range of nouns and hence are known as general classifiers (Gao & Malt, 2009).

For the purpose of noun prediction, nominal classifiers in Mandarin are similar to gender-marked determiners found in languages such as Dutch and Spanish in that, just like prediction-inconsistent gender-marked determiners, prediction-inconsistent classifiers can also signal to comprehenders that their noun prediction is likely incorrect. This is supported by recent ERP evidence reported by Kwon et al. (2017), who adapted the pre-target mismatch paradigm to Mandarin Chinese by using prediction-consistent vs. prediction-inconsistent classifiers in place of gender-marked determiners. They found that Mandarin Chinese speakers showed a significantly larger N400 response to prediction-inconsistent classifiers than to prediction-consistent classifiers, which suggests that Mandarin Chinese speakers are immediately sensitive to the mismatch between an incoming classifier and a predicted (and yet to appear) noun.

Further, previous studies have also shown that native Chinese speakers can use nominal classifiers incrementally to identify upcoming referents (Grüter, Lau & Ling, 2018; Huettig, Chen, Bowerman, & Majid, 2010; Klein, Carlson, Li, Jaeger & Tanenhaus, 2012; Lau & Grüter, 2015; Tsang & Chambers, 2011). For example, a visual world eye-tracking study by Huettig and colleagues (2010) showed that, upon hearing an utterance containing a specific classifier, Mandarin Chinese speakers were more likely to direct their eye gaze towards the objects that match the classifier than non-classifier-match objects. Similarly, in another visual world eye-tracking study, Klein and colleagues (2012) found that listeners were able to identify the target referent more quickly when the noun was preceded by a specific classifier than when it was preceded by a general classifier.

Taken together, the above findings suggest that listeners can (i) immediately detect a mismatch between a classifier and an expected noun, and (ii) use classifiers incrementally to anticipate upcoming nouns. In the present study we will build on these findings to ask whether, upon detecting a mismatch between a classifier and an expected noun, comprehenders can use the prediction-inconsistent classifier to immediately revise their noun prediction.

### ***The present study***

The design of the present experiment has three key elements. First, two experimental conditions were modelled on the pre-target mismatch paradigm used by Van Berkum et al. (2005). We constructed sentences with two plausible endings, one of which was highly predictable given the discourse context. For example, in a sentence like “*Jane went to Starbucks to buy ...*,” the most predictable noun was ‘coffee’. In one version, the sentence was continued with this noun which was preceded by a compatible specific classifier (e.g., ‘*bēi*’, the equivalent of ‘a cup of’ in English). In the other version, the sentence was

continued with a less expected but still contextually suitable noun such as ‘cake’. Crucially, this unexpected noun was selected from a distinct noun class than the expected noun, and as such the specific classifier that preceded the unexpected noun (e.g., ‘*kuài*’, the equivalent of ‘a piece of’ in English) was incompatible with the expected noun and vice versa.

Second, in order to examine how listeners’ expectations may be impacted by the specific classifiers, listeners’ eye-movements in the experimental conditions were compared against control conditions in which the specific classifiers were substituted with a general classifier that is compatible with both the expected and unexpected nouns. As such, the general classifier was uninformative about the noun that followed and served as a neutral baseline against which the specific classifier conditions were compared.

Lastly, in order to observe how the classifier may impact comprehenders’ predictions about the noun prior to its onset, we extended the time interval between the onset of the classifier and that of the noun by inserting a 4-to-5-syllable adjective that was compatible with both the expected and unexpected nouns between the classifier and the noun in all experimental sentences.

Based on previous findings, we expect that listeners can use the wider discourse to anticipate upcoming referents on the fly and as such should be more likely to look at the expected object (e.g., coffee in “*Jane went to Starbucks to buy ...*”) than the other objects before the onset of the classifier. Further, if and when listeners detect a mismatch between the classifier and the predicted noun, we should observe a decrease in looks to the expected object when they encounter a specific classifier that is not compatible with the expected noun (e.g., ‘*a piece of*’) relative to when the classifier is not informative (i.e., when they encounter a general classifier). Moreover, if and when listeners use the prediction-inconsistent classifier to revise their noun predictions, they should show an increase in looks to the unexpected

object that is compatible with the prediction-inconsistent classifier (e.g., cake) before the noun onset. Alternatively, if comprehenders cannot immediately use the specific classifier to revise their noun predictions, then they might not redirect their eye gaze to the unexpected object until after the noun onset.

## **Methods**

### ***Participants***

Twenty-three students (15 female and 8 male, mean age = 22.9 years, range 20 to 26 years) from University College London took part in the present study. All participants were native speakers of Mandarin from mainland China. All participants had normal hearing, normal or corrected-to-normal vision, and no known neurological disorder. All gave informed consent and were paid 7.5 pounds/hour for their participation. Data from seven additional participants were excluded due to either poor performances in comprehension questions in the filler items (accuracy < 90%) or excessive eye-movement artefacts such as blinks (see data processing procedures below).

### ***Materials***

Experimental stimuli consisted of 60 sets of sentences paired with black-and-white line drawings. A sample set of materials is shown in (3). We manipulated the predictability of the target noun in the discourse context (expected vs. unexpected noun) and the classifier that preceded the target noun (specific vs. general classifier), resulting in a  $2 \times 2$  design. The factor *noun expectancy* refers to the predictability of the target noun prior to the occurrence of the classifier, such that a target noun was considered “expected” or “unexpected” based on its predictability in the given discourse context alone and with no reference to the classifier. Further, the target nouns were always preceded by a classifier (general or specific) and an

adjective. In each item, the general classifier (e.g., ‘*xie*’) was compatible with both of the target nouns, but the specific classifier preceding the expected noun was incompatible with the unexpected noun (e.g., the classifier ‘*duō*’ is incompatible with *leaf*) and the specific classifier preceding the unexpected noun was incompatible with the expected noun (e.g., the classifier ‘*piàn*’ is incompatible with *flower*). Lastly, all experimental items were constructed in pairs such that the expected noun in one item served as the unexpected noun in the other item and vice versa. This allowed us to compare an identical set of classifiers and target nouns across the expected and unexpected conditions.

(3) 在花园里玩耍时，小男孩送给小女孩...  
 at garden play when little boy give little girl ...

- a. *Expected noun, specific classifier:*  
 一朵很漂亮的花  
 one CL<sub>duō</sub> very beautiful flower
- b. *Expected noun, general classifier:*  
 一些很漂亮的花  
 one CL<sub>xie</sub> very beautiful flower
- c. *Unexpected noun, specific classifier:*  
 一片很漂亮树叶  
 one CL<sub>piàn</sub> very beautiful leaf
- d. *Unexpected noun, general classifier:*  
 一些很漂亮树叶  
 one CL<sub>xie</sub> very beautiful leaf

“*While playing in the garden, the little boy gave the little girl a very beautiful flower / leaf.*”

The experimental stimuli were developed in two stages. Initially we created a total of 80 pairs (160 sets) of sentence frames. To determine the predictability of different nouns in these sentence frames, we conducted a cloze probability norming study using the online platform Ibxfarm (Drumond, 2013). Each sentence frame had three versions, two ending with a specific classifier and one with a general classifier. Sentence frames were divided into 3

presentation lists, such that each list contained exactly one version of each item. We collected cloze data from 90 native Mandarin Chinese speakers from South China Normal University (30 participants per presentation list) who did not participate in the eye-tracking experiment. In accordance with standard cloze norming procedures, participants were asked to read each sentence frame and to supply a word or phrase that they expected to see next. In almost all trials participants responded with a noun. The cloze probability of a word in a given context is defined as the proportion of trials on which speakers continue the sentence frame with that word (Taylor, 1953).

From the resulting database, 30 pairs (that is, 60 sets) of items and their target nouns were selected for use in the eye-tracking experiment. All selected items fulfilled the following criteria: (i) both target nouns were compatible with the general classifier, (ii) the expected target noun had a higher cloze probability than the unexpected target noun in the general classifier conditions, (iii) the expected target noun was incompatible with the specific classifier in the unexpected condition, (iv) the unexpected target noun was incompatible with the specific classifier in the expected condition, and (v) the unexpected target noun had a higher cloze probability in the specific classifier condition than in the general classifier condition. On average, the expected and unexpected target nouns in the general classifier conditions had a cloze probability of 52% (range = 10% to 96%) and 5% (range = 0% to 26%) respectively. In the specific classifier conditions, the same expected and unexpected target nouns had an average cloze probability of 80% (range = 31% to 100%) and 60% (range = 13% to 100%) respectively. Subsequently, a 4-to-5-syllable adjective which was compatible with both target nouns was inserted between the classifier and the target noun in all experimental sentences, making the stimulus onset asynchrony (SOA) between the classifier and the noun 1217 ms on average. A complete list of the experimental sentences is available in the Supplementary Materials.

Each item was paired with a visual display of black-and-white line drawings depicting the target nouns and two distractor objects. All drawings were sourced from the Noun Project (<https://thenounproject.com>). As the experimental items were constructed in pairs, each drawing that depicted a target noun appeared once as the expected object and once as the unexpected object. In order to ensure that it was not possible for participants to identify the target object in a given trial based on their familiarity of the drawings alone, in some cases a target object in an experimental item was used as a distractor in another item, and a majority of non-target objects (i.e., distractor objects in the experimental items and objects in the filler items) were also presented in more than one trial. A sample visual display is shown in Figure 1. The nouns depicted by the distractor objects were phonologically dissimilar to the target nouns and were not compatible with either specific classifier in that item. Further, we also avoided using distractor objects that were visually similar to the target objects. This is because previous studies have shown that listeners were more likely to look at distractors that share overlapping phonological or visual features with the targets than unrelated distractors (Hintz & Huettig, 2015; Huettig & McQueen, 2007). The position of the drawings was pseudo-randomised to ensure that the target object appeared in all four positions with equal probability.



Figure 1. Sample visual display. Credits (clockwise from top left): 'Leaf' icon created by Ozza Okuonghae, 'bird' icon by Sebastian Andreasson, 'flower' icon by Adam Zubin and 'candy' icon by Petra Prgomet from the Noun Project ([thenounproject.com](https://thenounproject.com)).

The 60 sets of experimental sentences were divided into 4 presentation lists, such that each list contained exactly one version of each item. An additional 30 unrelated filler sentences with similar length and structural complexity were included in all presentation lists. Among the filler items, a nominal classifier was present in nine of them; three of them were compatible with more than one of the objects in the visual display. All filler sentences were followed by a simple Yes/No comprehension question to ensure that participants pay attention to the stimuli; experimental sentences were not followed by a comprehension question to avoid directing participants' attention to the experimental manipulations. As a result, a comprehension question was presented in one third of the trials. All sentences were recorded by the second author – a female native speaker of Mandarin Chinese – with a normal speech rate and intonation (e.g., no stress on either the classifier or the target noun). The order of sentences in each list was pseudo-randomised; care was taken to ensure that experimental sentences from the same pair (that is, two items with the same target nouns in different conditions) were never presented consecutively.

### ***Procedure***

Participants were seated at a comfortable distance (about 60 cm) from a 23" computer screen in a testing room. A Tobii TX300 was used to monitor participants' eye-movements with a sampling rate of 300 Hz while they listened to the spoken stimuli through headphones. A 9-point calibration procedure was performed at the beginning of each testing session. Prior to each trial, a fixation cross appeared at the centre of the screen to allow automatic correction for drifts. An array of four line drawings appeared on the screen 1500 ms prior to the onset of the spoken stimuli and stayed on the screen until the end of a trial. Each testing session was divided into three blocks of 30 trials each. Participants were instructed to listen to the sentences carefully and to answer comprehension questions by pressing one of two buttons

on the keyboard. Before the experimental session, participants completed four practice trials to familiarise themselves with the task. Including set-up time, each session lasted approximately 45 minutes.

### ***Data processing and analysis***

Eye-movement data from 1000ms before the onset of the classifier up to the end of a trial (or 4000 ms after classifier onset, whichever is earlier) were extracted for analysis. Trials with more than 50% track loss during this interval were excluded from analysis, leaving an average of 44 trials per participant. The average number of trials excluded due to track loss was evenly distributed across conditions (range = 3.7 to 4.2).

Each eye-tracking data sample was scored as '1' or '0' depending on whether the target was fixated on in that sample (samples with track loss were excluded). The target was the object ultimately named in the sentence, and thus it was the expected object (e.g., 'flower' in (3)) in the expected N conditions and the unexpected object (e.g., 'leaf' in (3)) in the unexpected N conditions. We time-locked participants' eye-tracking data in each trial to the onset of the classifier. Following Altmann (2011), we did not offset the data in time, such that time zero always marked the actual onset of the classifier.

For the time-window analysis, we computed grouped binomial data in five time windows for each trial using the number of samples in which the target was fixated on (and the number of samples in which it was not). Three time windows were time-locked to onset and offset of the *classifier*, *adjective*, and *noun* respectively. In addition, the *pre-classifier* window delimits the 1000ms prior to the onset of the classifier, and the *post-noun* window delimits the time between the offset of the noun and the end of trial.

We analysed the grouped binomial eye-tracking data in each time window using mixed-effects logistic regression (Jaeger, 2008; Donnelly & Verkuilen, 2017). We used the `glmer` function from the *lme4* R package to fit the models (version 1.1-12; Bates, Maechler, Bolker & Walker, 2015) in R (R Development Core Team, 2015). All models had noun expectancy and classifier type and their interaction as fixed effects, with participants and items as random effects. Sum-coding was used for both noun expectancy and classifier type. In addition, the maximal random effects structure (random intercepts and slopes by participants and by items) was used in all models (Barr, 2008).<sup>3</sup> The analysis yields regression coefficients ( $\beta$ ), along with their standard errors and Wald's *Z*-score.

Further, in order to examine the time course of the effect of classifier type on listeners' eye-movements in the unexpected noun conditions, we conducted a cluster-based permutation analysis (Maris & Oostenveld, 2007) using the R package *eyetrackingR* (Dink & Ferguson, 2015). Our goal was to identify cluster(s) in the data that arose from listeners directing their attention (i) away from the expected object and (ii) to the unexpected object following an unexpected specific classifier (relative to the general classifier control). We analysed the data starting from 1000 ms prior to classifier onset to the end of trial in 20 ms time bins to examine when the two unexpected conditions diverged from each other. Listeners' looks to the expected object and their looks to the unexpected (target) object were analysed separately.

This analysis involves two stages. In the cluster formation stage, a two-tailed t-test comparing the two unexpected conditions was conducted on data in each 20 ms time bin to identify

---

<sup>3</sup> The model for data in a given time window was specified as: `glmer(cbind(SamplesInAOI, SamplesTotal-SamplesInAOI) ~ Noun * Classifier + (1 + Noun * Classifier | Item) + (1 + Noun* Classifier | Subject), data = data, family = binomial)`

clusters of time bins in which the threshold statistic was exceeded ( $\alpha = .05$ ). A cluster statistic was computed by summing the statistics for the time bins within a cluster. Subsequently, in the permutation and inference stage, the data were shuffled<sup>4</sup>, the same statistical test was conducted to identify new clusters, and the largest summed-statistic from each iteration was obtained. This process was iterated 2,000 times to create a distribution of summed-statistics (a null distribution), against which the cluster statistic was compared to obtain a  $p$  value.

The cluster-based permutation analysis assumes statistical independence between trials and is preferred over testing time bins independently because it controls the family-wise error rate. However, this analysis is not equipped to provide precise estimates of the onset of an effect (that is, the time point at which two conditions begin to diverge). This is because the extent of the cluster(s) in the data is determined at the initial stage and is not the subject of an inferential test with guaranteed error rates. In other words, what passes the threshold at the initial stage depends on the amount of noise in the data, and with more trials (or less noise) a cluster may have an earlier onset and a later offset. For further discussion and simulation data, see Sassenhagen & Draschkow (2019).

---

<sup>4</sup> Shuffling involves combining trials from both conditions and randomly drawing trials from the combined data set to form two new subsets.

## Results

### *Time-window analysis*

The proportions of fixations on the target object (the object ultimately named in a given trial) across conditions are shown in Figure 2. Results of the logistic mixed effect models are presented in Table 1. Full model outputs are available in the Supplementary Materials.

In both the pre-classifier and classifier windows, logistic mixed models revealed a significant effect of noun expectancy (pre-classifier:  $\beta = 0.29$ ,  $SE = 0.08$ ,  $Z = 3.55$ ,  $p < .01$ ; classifier:  $\beta = 0.46$ ,  $SE = 0.08$ ,  $Z = 5.32$ ,  $p < .01$ ). As the target object was the expected object (e.g., flower) in the expected noun conditions and the unexpected object (e.g., leaf) in the unexpected noun conditions, this shows that listeners were already more likely to be looking at the expected object than the unexpected object before encountering the classifier.

In the adjective window, there was a significant effect of both noun expectancy ( $\beta = 0.31$ ,  $SE = 0.08$ ,  $Z = 4.10$ ,  $p < .01$ ) and classifier type ( $\beta = 0.27$ ,  $SE = 0.06$ ,  $Z = 4.69$ ,  $p < .01$ ). The observation of an effect of classifier type in this interval shows that listeners became more likely to fixate on the target object (that is, the object that was going to be named in that trial, regardless of whether it was initially expected or not) when the classifier was informative (specific classifier) than when it was not (general classifier).

In the noun window, we observed main effects of noun expectancy ( $\beta = 0.26$ ,  $SE = 0.06$ ,  $Z = 4.54$ ,  $p < .01$ ) and classifier type ( $\beta = 0.31$ ,  $SE = 0.07$ ,  $Z = 4.63$ ,  $p < .01$ ) along with a significant interaction effect ( $\beta = -0.24$ ,  $SE = 0.11$ ,  $Z = -2.06$ ,  $p < .05$ ). This shows that the effect of noun expectancy (i.e., the difference between the expected and unexpected noun conditions) was greatly reduced in the specific classifier condition compared to the general classifier condition.

No effects were observed in the post-noun window, as listeners' looks to the target object reached ceiling in all four conditions.

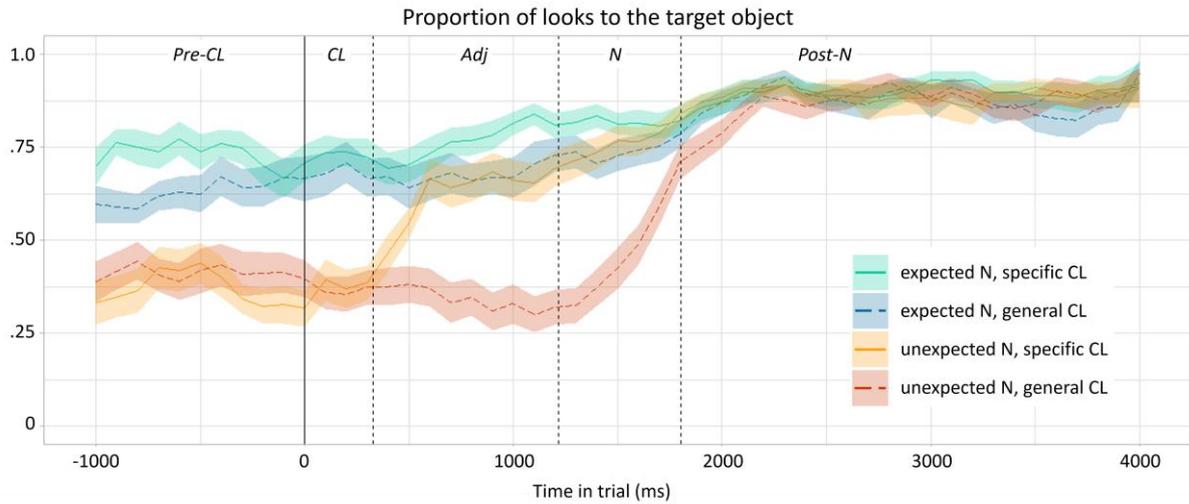


Figure 2. Proportion of looks to the target object time-locked to the onset of the classifier ( $t = 0$  ms) across conditions. The target object was the expected object (flower) in the expected N conditions and the unexpected object (leaf) in the unexpected N conditions. The dashed vertical lines indicate the mean offset of the classifier, adjective, and noun respectively. Standard errors are shown using partially transparent shading.

Table 1. Logistic mixed effect model results (coefficients, standard errors and Wald's Z-scores) for all time windows.

	Pre-Classifier			Classifier			Adjective			Noun			Post-Noun		
	$\beta$	SE	Z	$\beta$	SE	Z	$\beta$	SE	Z	$\beta$	SE	Z	$\beta$	SE	Z
(Intercept)	0.02	0.03	0.74	0.06	0.04	1.52	0.17	0.03	5.16**	0.27	0.03	8.42**	0.54	0.02	21.70**
Noun	0.29	0.08	3.55**	0.46	0.08	5.62**	0.31	0.08	4.10**	0.26	0.06	4.54**	-0.01	0.03	-0.16
Classifier	0.02	0.07	0.27	0.02	0.08	0.26	0.27	0.06	4.69**	0.31	0.07	4.63**	0.01	0.03	0.34
Noun:Classifier	0.17	0.11	1.49	0.09	0.13	0.68	-0.19	0.12	-1.60	-0.24	0.11	-2.06*	0.03	0.06	0.39

\*  $p < .05$ ; \*\*  $p < .01$

### Time course analysis

The time course for the effect of classifier type on listeners' looks to the expected and unexpected objects in the unexpected noun conditions are shown in Figure 3.

The cluster-based permutation analysis on listeners' looks to the expected object in the unexpected noun conditions revealed a significant effect of classifier type ( $p < .0001$ ). This corresponded to a cluster in the observed data from 460 to 2080 ms after the onset of the

classifier. Given that the onset of the noun occurred at least 845 ms after the classifier onset (with an average of 1217 ms), this shows that listeners' looks to the expected object were significantly reduced in the specific classifier condition compared to the general classifier condition prior to the noun onset.

This finding was mirrored by the analysis of listeners' looks to the unexpected (target) object, which also revealed a cluster ( $p < .0001$ ) from 460 to 2080 ms after the onset of the classifier. This suggests that listeners' reduced looks to the expected object completely coincided with their increased looks to the unexpected (target) object.



Figure 3. Time course for the effect of classifier type on listeners' looks to the expected object (top panel) and unexpected object (bottom panel) in the unexpected noun conditions. Standard errors are shown using partially transparent shading. Intervals with a statistically significant effect based on the cluster-based permutation analysis are marked with an asterisk.

## Discussion

In the present study we set out to investigate what happens when comprehenders encounter cues in the input that are not consistent with the existing predictions. We took advantage of the rich nominal classifier system in Mandarin Chinese to ask whether and how listeners use prediction-inconsistent classifiers to revise their existing predictions for an upcoming noun.

This study has three main findings. First, prior to the classifier onset, listeners were already more likely to look at the object depicting the most expected noun than other objects. This suggests that, even prior to the classifier, listeners had already formed predictions about the upcoming noun and that they were anticipating the expected noun. This is consistent with a large body of evidence showing that listeners can rapidly integrate rich contextual information to anticipate upcoming referents (Kamide, 2008; Kutas et al., 2011).

Second, we found that listeners began looking away from the originally expected object within a few hundred milliseconds of hearing a classifier that was incompatible with the expected noun. Taking into consideration the time it takes to programme a saccade (~200ms; Allopenna, Magnuson, & Tanenhaus, 1998), this finding shows that listeners were immediately sensitive to the incompatibility between the classifier and the noun they expected. This is in line with Kwon et al.'s (2017) observation that prenominal classifiers elicited a larger N400 response when they were inconsistent with an expected noun, and more generally, the observation that prediction-inconsistent prenominal elements elicited a differential ERP response in studies across different languages (e.g., Van Berkum et al., 2005, Wicha et al., 2004).

Third, and more importantly, the time course analysis revealed that listeners quickly began to redirect their eye-gaze from the expected object to the (initially unexpected) target object

upon encountering an unexpected specific classifier. This suggests that listeners were able to use the information conveyed by the specific classifier in conjunction with the preceding context to revise their noun predictions and to direct their eye-gaze onto the target. Taken together, the present results show that, upon encountering a prediction-inconsistent classifier, listeners were able to detect a mismatch between the classifier and the predicted noun, and immediately use the unexpected classifier to revise their noun predictions without waiting for the noun to appear in the bottom-up input.

### ***What did listeners do to revise their predictions?***

One question that arose, however, is the extent to which listeners could have identified the target without fully using the information conveyed by the classifier. For instance, if the expected and unexpected objects were the only elements in the visual display that were compatible with the sentence context leading up to the classifier, then the identity of the target would have been apparent as soon as the expected object has been ruled out by the prediction-inconsistent classifier. In such a scenario, ruling out the expected object leaves only one viable option, so listeners need not rely on the classifier to identify a compatible object.

In order to examine this further, we evaluated the visual display for each experimental item to determine whether the target object was the only contextually appropriate alternative to the expected object in each case. We found that a total of 45 items (out of 60 experimental items) contained at least one distractor object in the visual display (in addition to the expected and unexpected objects) that was considered compatible with the sentence context leading up to

the classifier<sup>5</sup>. In other words, in these items, listeners were faced with at least two contextually appropriate objects even after having ruled out the expected object and must use the classifier to identify the target object.

We redid the cluster-based permutation analyses (see Analysis) on the data from these 45 items to identify cluster(s) in the data that arose from differences between the two unexpected noun conditions, both in terms of listeners' looks to the expected object as well as their looks to the (initially unexpected) target object. Once again we found a significant effect of classifier type in both cases ( $p < .0001$ ). This corresponded to a cluster in the observed data between 520 and 1880 ms after the classifier onset. Compared to the original analyses, the cluster identified here had both a slightly later onset (520 ms vs. 460 ms post-classifier onset) and an earlier offset (1880ms vs. 2080ms). This is likely due to the fact that the present analysis included ~25% fewer trials than the original analysis and thus had less statistical power to detect differences, especially at the edges of the cluster where the effect was smaller (see Figure 2). More importantly, since this effect clearly preceded the noun onset (which was at least 845 ms after the classifier onset), these results show that, upon encountering an unexpected classifier, listeners were not only able to immediately rule out the expected object, but they were also able to use the classifier to identify the target object among the alternatives before the noun appeared in the input.

---

<sup>5</sup> These items were selected based on the judgements of 5 native speakers who did not participate in the eye-tracking study. They were given the sentence contexts leading up to the target noun phrase (e.g., "While playing in the garden, the little boy gave the little girl ...") and asked to judge whether the sentence can be continued with either of the distractor objects. The average acceptability rating (that at least one of the distractor objects was compatible with the sentence context) was 75% for the 45 selected items.

### *Speed of prediction revision*

In the present study, we observed no delay between the time at which listeners looked away from the expected object and when they started looking towards the target object (that is, the object that was ultimately named) in either analysis. This is somewhat unexpected, as it suggests that listeners did not have to search for an alternative among the objects at all when the expected object was ruled out by the prediction-inconsistent classifier. One possible interpretation is that listeners were able to revise their prediction instantaneously once an existing prediction had been disconfirmed. Alternatively, listeners' identification of the correct target may have been aided by the fact that there were only a small number of objects displayed on the screen which effectively restricted the set of candidates to just a handful. As the objects were present from the beginning of a trial, listeners were likely able to identify them and encode their locations on the screen early on, which may explain why they can redirect their eye-gaze to the correct target as soon as the expected object was ruled out.

The visual world paradigm in the present study has allowed us to ask *whether* listeners can use a prediction-inconsistent cue (an unexpected classifier) to revise their predictions for an upcoming word. The answer is yes. However, since the visual display effectively restricted the set of candidates to a handful for the listeners, questions remain about how listeners may revise their predictions when they are not aided by a restricted visual context. Further research will be needed to investigate how listeners' capacity to revise their predictions on the fly may be impacted by the visual context.

### **Conclusion**

The present study examined whether and how comprehenders may revise their existing predictions upon encountering a prediction-inconsistent cue. By capitalising on the rich

classifier system in Mandarin Chinese, we showed that listeners were not only able to quickly detect a mismatch between an incoming classifier and an expected (but yet to appear) noun, but they also were able to use the prediction-inconsistent classifier to identify the correct object. This study provides initial evidence for comprehenders' capacity to quickly use a prediction-inconsistent cue to revise their predictions for upcoming inputs.

## **Acknowledgements**

This work was supported by a British Academy/ Leverhulme Trust Small Research Grant (SRG000032444). We would like to thank Suiping Wang for help with norming data collection.

## **References**

- Allan, K. (1977). Classifiers. *Language* 53: 285-311
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(38), 419–439.  
<https://doi.org/10.1006/jmla.1997.2558>
- Altmann, G. T. M. (2011). Language can mediate eye movement control within 100milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, 137(2), 190–200. <https://doi.org/10.1016/j.actpsy.2010.09.009>
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.  
[https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.  
<https://doi.org/10.1016/j.jml.2007.09.002>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bendixen, A., Schroger, E., & Winkler, I. (2009). I Heard That Coming: Event-Related

- Potential Evidence for Stimulus-Driven Prediction in the Auditory System. *Journal of Neuroscience*, 29(26), 8447–8451. <https://doi.org/10.1523/jneurosci.1493-09.2009>
- Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 15, 607–624. <https://doi.org/10.3758/s13415-015-0340-0>
- Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, 33, 803–828. <https://doi.org/10.1080/23273798.2018.1427878>
- Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2016). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, 31(5). <https://doi.org/10.1080/23273798.2015.1066832>
- Clifton, C. J., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49(3), 317–334. [https://doi.org/10.1016/S0749-596X\(03\)00070-6](https://doi.org/10.1016/S0749-596X(03)00070-6)
- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61(1), 150–162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>
- 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. <https://doi.org/10.1038/nn1504>
- Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R Library for Eye-tracking Data Analysis*. Retrieved from <http://www.eyetrackingr.com>
- Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language*, 94, 28–42. <https://doi.org/10.1016/j.jml.2016.10.005>
- Drummond, A. (2013). IbeXFarm (Version 0.3) [Software]. Available online at: <http://spellout.net/ibexfarm/>
- Erbaugh, M.S. (2004). Chinese classifiers: Their use and acquisition. In P. Li, L.H. Tan, E. Bates and O.J.L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 39-51), Cambridge, UK: Cambridge University Press
- 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. <https://doi.org/10.1006/jmla.1999.2660>

- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*(1), 75–84. <https://doi.org/10.1016/j.brainres.2006.06.101>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200–214. <https://doi.org/10.1016/j.jml.2017.04.007>
- Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes, 24*, 1124–1179. <https://doi.org/10.1080/01690960802018323>
- Grüter, T., Lau, E., & Ling, W. (2018). L2 Listeners Rely on the Semantics of Classifiers to Predict. In A. B. Bertolini & M. J. Kaplan (Eds.), *BUCLD 42: Proceedings of the 42nd annual Boston University Conference on Language Development* (pp. 303–316). Cascadilla Press.
- Grüter, T., Rohde, H., & Schafer, A. J. (2015). Discourse Expectations in a Non-native Language. In *DETEC2015*.
- Hintz, F., & Huettig, F. (2015). The Complexity of the Visual Environment Modulates Language-Mediated Eye Gaze. In R. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and Vision in Language Processing* (pp. 39–55). Berlin: Springer. <https://doi.org/10.1007/978-81-322-2443-3>
- Huettig, 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. <https://doi.org/10.1016/j.jml.2007.02.001>
- Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture, 10*, 39–58.
- 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. <https://doi.org/10.1016/j.jml.2015.10.007>
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience, 32*(8), 954–965.

<https://doi.org/10.1080/23273798.2016.1242761>

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.

Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Linguistics and Language Compass*, 2(4), 647–670. <https://doi.org/10.1111/j.1749-818X.2008.00072.x>

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. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)

Klein, N. M., Carlson, G. N., Li, R., Jaeger, T. F., & Tanenhaus, M. K. (2012). Classifying and massifying incrementally in Chinese language comprehension. In D. Massam (Ed.), *Count and Mass Across Languages* (pp. 261–282). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199654277.003.0014>

Kochari, A. R., & Flecken, M. (2018). Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*. <https://doi.org/10.1080/23273798.2018.1524500>

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8(4), 533–572. <https://doi.org/10.1080/01690969308407587>

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

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*. <https://doi.org/10.1093/acprof:oso/9780195395518.003.0065>

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, 166, 433–446. <https://doi.org/10.1016/j.cognition.2017.06.010>

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, 166, 433–446. <https://doi.org/10.1016/j.cognition.2017.06.010>

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933. <https://doi.org/10.1038/nrn2532>

- Lau, E., & Grüter, T. (2015). Real-time processing of classifier information by L2 speakers of Chinese. *Proceedings of the 39th Annual Boston University Conference on Language Development*, 311–323.
- Li, C.N., & Thompson, S.A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Nieuwland, M. S., Politzer-ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... Kohu, Z. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, 7, e33468. <https://doi.org/10.7554/eLife.33468>
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464–496. doi:10.1080/01638530802356463
- Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, 1291, 92–101. doi:10.1016/j.brainres.2009.07.042
- Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, 8(1), 89. doi:10.1186/1471-2202-8-89
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), 1–8. <https://doi.org/10.1111/psyp.13335>
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2), 109–147. [https://doi.org/10.1016/S0010-0277\(99\)00025-6](https://doi.org/10.1016/S0010-0277(99)00025-6)
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13, 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Szewczyk, J. M. (2018). *Prediction-inconsistent information leads to prediction updating – an ERP study on sentence comprehension*. Talk given at the 31st annual CUNY sentence processing conference, Davis, CA, USA.

- Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. <https://doi.org/10.1016/j.jml.2012.12.002>
- Szewczyk, J. M., & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*, 33(6), 665–686. <https://doi.org/10.1080/23273798.2017.1401101>
- Tai, J. H.-Y (1994). Chinese classifier systems and human categorization. In M. Y. Chen & O. J. L. Tzeng (Eds.). *In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change* (pp. 479–494). Taipei: Pyramid Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. <https://doi.org/10.1126/science.7777863>
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30(4), 415.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*. <https://doi.org/10.1006/jmla.1994.1014>
- Tsang, C., & Chambers, C. G. (2011). Appearances Aren't Everything: Shape Classifiers and Referential Processing in Cantonese. *Journal of Experimental Psychology: Learning Memory and Cognition*, 37(5), 1065–1080. <https://doi.org/10.1037/a0023601>
- 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. <https://doi.org/10.1037/0278-7393.31.3.443>
- van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>
- Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346(3), 165–168. doi:10.1016/S0304-3940(03)00599-8
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in spanish. *Cortex*, 39(3), 483–508. doi:10.1016/S0010-9452(08)70260-0

- 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. <https://doi.org/10.1162/0898929041920487>
- Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356–366. <https://doi.org/10.1016/j.neuroimage.2012.04.054>
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*, 729–732.