

## Prediction in the maze: Evidence for probabilistic pre-activation from the English *a/an* contrast

E. Matthew Husband, University of Oxford, UK, [matthew.husband@ling-phil.ox.ac.uk](mailto:matthew.husband@ling-phil.ox.ac.uk)

---

The idea that comprehenders predict upcoming linguistic content has become core to many theories of language processing. Experimental studies exploiting morphosyntactic and phonotactic constraints on a word form preceding a high cloze target word have been key to underpinning predictive accounts of comprehension, but investigating these tight sequential contrasts with traditional behavioral methods is difficult. The *maze task*, with its more focal measure of incremental processing, may provide a cheap and easy methodology to study early cues to prediction. An experiment investigating the *a/an* contrast (DeLong, Urbach, & Kutas, 2005; Nieuwland, et al., 2018) using A-maze (Boyce, Futrell, & Levy, 2020) finds that unexpected articles, as well as nouns, elicit slower focal response times. Response times are also shown to be inversely related to noun cloze probabilities, with slower responders showing larger effects of expectation. This study demonstrates that the maze task can be sensitive to expectation and is a useful alternative methodology for investigating prediction in comprehension.

---



## Introduction

The past two decades has seen a resurgence of interest in the idea that comprehenders routinely engage in prediction of upcoming linguistic content. Prediction is now often assumed to occur along all levels of linguistic representation, from semantics and morphosyntax to phonology/orthography, serving to facilitate the access and integration of bottom-up information into unfolding sentential and discourse representations.

A key observation supporting theories that comprehenders predict upcoming words comes from experimental research that exploits morphosyntactic and phonotactic constraints on the form of words that precede high cloze target words. DeLong, Urbach, and Kutas (2005) uses the phonotactic constraints of English indefinite articles to investigate probabilistic pre-activation of expected (high-cloze) nouns. They presented participants with highly constraining contexts like “*The day was breezy so the boy went outside to fly...*” and manipulated the article and noun so that the noun was either expected, *a kite*, or unexpected, *an airplane*. The amplitude of the N400 elicited by the preceding *a/an* articles was found to decrease as the cloze probability of nouns increased. Other ERP studies investigating morphosyntactic constraints (Foucart, et al., 2014; Martin, et al., 2018; Otten & Van Berkum, 2009; Wicha, Moreno, & Kutas, 2004), phonotactic constraints (Martin, et al., 2013), or both (Ito, et al., 2020) report evidence for pre-activation of target nouns on preceding articles. Although some studies have not successfully replicated these effects (Ito, Martin, & Nieuwland, 2017; Nieuwland, Arkhipova, & Rodríguez-Gómez, 2020; Nieuwland, et al., 2018), taken together this research has underpinned theories of prediction in language comprehension by providing unambiguous evidence for predictive processes.

Critical to the success of some of these studies is the high temporal resolution and word-by-word presentation typical of ERP methodology in reading. This is particularly true for phonotactic/orthographic constraints which occur between adjacent words. These tight sequential constraints prevent researchers from investigating these phenomena with traditional incremental reading tasks. Self-paced reading is one of the most commonly used techniques and is relatively cheap, especially with recent innovations in online participant crowdsourcing and web-based stimulus delivery. However, it is difficult to isolate effects on a specific word due to well-known spillover effects which smear processing differences across multiple words (Mitchell, 1984; Witzel, Witzel, & Forster, 2012). Eye movements in reading are also a poor measure. Short functional words are often skipped during reading and are likely available in parafoveal preview, requiring very careful experimental manipulation to isolate effects on them (Balota, Pollatsek, & Rayner, 1985; Cutter, Martin, & Sturt, 2020).

Research on prediction has been held back by the lack of a cheap and easy behavioral methodology to study early cues to prediction error (cf. Van Berkum, et al., 2005). The maze task, however, offers an alternative incremental reading method that can address some of the potential shortcomings of the methods mentioned above (Forster, Guerrera, & Elliot, 2009). In maze tasks,

sentences are presented to participants as a sequence of choices between two alternatives. One alternative is the correct continuation of the sentence while the other is a distractor. Distractors are real words that are anomalous given current sentence context (G(rammaticality)-maze) or pseudowords (L(exicality)-maze). Maze tasks have been shown to deliver focal measures of processing difficulty, with effects occurring on their trigger word, that are comparable to self-paced reading and eye tracking (Witzel, Witzel, & Forster, 2012; Witzel & Forster, 2014).

Maze tasks have been difficult to implement because pairing a distractor word to each word in all sentence material was labor-intensive and prone to researcher error. Addressing these difficulties, Boyce, Futrell, and Levy (2020) used natural language processing to automate distractor selection, creating the A(uto)-maze. They demonstrated that the A-maze elicited focal reading time effects on the disambiguating word of three types of attachment ambiguities with online participants. A-Maze effects were comparable to G-maze and more sensitive than L-maze and self-paced reading.

Given that focal effects can be found with A-maze, A-maze response times might also isolate prediction error effects on expectation-mismatching article forms preceding unexpected nouns. As a participant-controlled continuous temporal measure, A-maze response times might also provide new information about predictive processing, e.g. via natural differences in participant's comprehension speed. To investigate these possibilities, an A-maze task was conducted on high cloze probability sentence contexts manipulating noun predictability and the form their immediately preceding articles.

## **Experiment Method**

*Participants.* 40 native UK English speakers (24 female, 14 male, 1 other; ages 18–71, avg.: 35, sd.: 14) were recruited on Prolific (Peer, et al., 2017) and paid £5 for participation. One participant's results did not transfer, leaving 39 participants for initial analysis.

*Items.* 80 sentence contexts were combined with two possible continuations, an expected and an unexpected indefinite article plus noun combination, from Nieuwland et al.'s (2018) replication of DeLong, Urbach, and Kutas (2005). Each article-noun combination appeared once as the expected continuation and the other as the unexpected continuation in different contexts, for 160 sentences total. Median cloze probability for expected articles was 0.75 (min 0.16; mean 0.74; max 1.00), for expected nouns 0.90 (min 0.23; mean 0.82; max 1.00), for unexpected articles 0.02 (min 0.00; mean 0.08; max 0.39), and for unexpected nouns 0.00 (min 0.00; mean 0.09; max 0.77). Expected article and noun cloze probability correlation was 0.24, and 0.10 for unexpected conditions. Sentences were divided in to two lists of 80 sentences each. Each article-noun combination appeared only once per list. A yes/no comprehension question followed 21 of the sentences.

Distractor words were generated using A-maze (Boyce, Futrell, & Levy 2020; <https://vboyce.github.io/Maze/>) with the Gulordava language model (Gulordava, et al., 2018). This process selected a distractor word for all words but the first of each sentence stimulus. Distractor words were matched in terms of length and approximate frequency to the correct continuation word and were low probability given the left sentence context of that word. The left/right position of correct and distractor words was randomized, except for the first word of each sentence where the correct word was presented on the left against a distractor “x-x-x”. Examples are given in **Table 1** and a full stimulus set including paired distractors and comprehension questions are available at <https://osf.io/frdtm/>.

**Table 1:** Example sentences and distractors generated by A-maze. Target articles and nouns are underlined as are their paired distractor words.

	Sentence
<b>Expected</b>	<p>The highlight of Jack’s trip to India was when he got to ride <u>an elephant</u> in the parade.</p> <p>x-x-x subjected wish Nuclei tons cent Ratio boys file miss skin mean inch <u>lie extends</u> pm knew trends.</p> <p>You never forget how to ride <u>a bicycle</u> once you’ve learned.</p> <p>x-x-x hours animal door fund onto <u>lack deposits</u> glad author eastern.</p>
<b>Unexpected</b>	<p>The highlight of Jack’s trip to India was when he got to ride <u>a bicycle</u> in the parade.</p> <p>x-x-x subjected wish Nuclei tons cent Ratio boys file miss skin mean inch <u>lie extends</u> pm knew trends.</p> <p>You never forget how to ride <u>an elephant</u> once you’ve learned.</p> <p>x-x-x hours animal door fund onto <u>lack deposits</u> glad author eastern.</p>

*Procedure.* Sentences were presented using the IbeXFarm web-based platform (<https://ibex.spellout.net/>) with Boyce, Futrell, and Levy’s (2020) Maze module controller in ‘redo’ mode. Participants used ‘e’ and ‘i’ keys to select the left or right alternative continuation, respectively. Selecting the correct continuation word advanced the sentence to the next word pair. Selecting the distractor word elicited an error message, “Incorrect! Please try again”, prompting the participant to select the correct continuation. ‘e’ and ‘i’ keys were also used to answer yes/no comprehension questions, presented in full on the screen to participants.

*Data Analysis.* Item 29 was removed due to a coding error. Regression models for dependent variables (Error Rates, Response Times) were constructed using either the categorical factor (Expectation, sum coded 0.5 [unexpected] and -0.5 [expected]) or graded by-item noun cloze probabilities (Noun Cloze, centered prior to model fit). Error rates were analyzed using Firth’s penalized likelihood method (Firth, 1993), using *logistf* (Heinze, Ploner, & Jiricka, 2020). This

analysis was considered more appropriate than mixed-effects logistic regression given the very low error rates.

Words with error responses (including post-error ‘correct’ responses) were removed and the remaining response times for all correct-on-first-attempt words were analyzed by fitting mixed-effects models with maximal random effects using *lme4* (Barr, et al., 2013) and *p*-values determined by *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) using the Satterthwaite approximation. Conditional means were derived from the models using *emmeans* (Lenth, 2021) and difference-adjusted 95% (percentile) mixed-effect-model-based intervals were calculated to account for crossed within-subjects and within-items random effects (Politzer-Ahles, 2017). Analyses are reported using untransformed RT, though  $\log(\text{RT})$  revealed similar patterns. Raw data and analysis scripts are available at <https://osf.io/frdtm/>. Additional analyses are reported in the appendix.

## Results

*Comprehension accuracy.* Overall accuracy was very high (90%). One participant whose accuracy was 2 standard deviations below the mean (52%) was removed from further analysis. All other participants had average accuracies above 70% and were retained for analysis, with 91% overall accuracy average.

*Reading measures.* The following analyses are restricted to the two critical regions (target article and noun) and the three words preceding and following the two critical words.

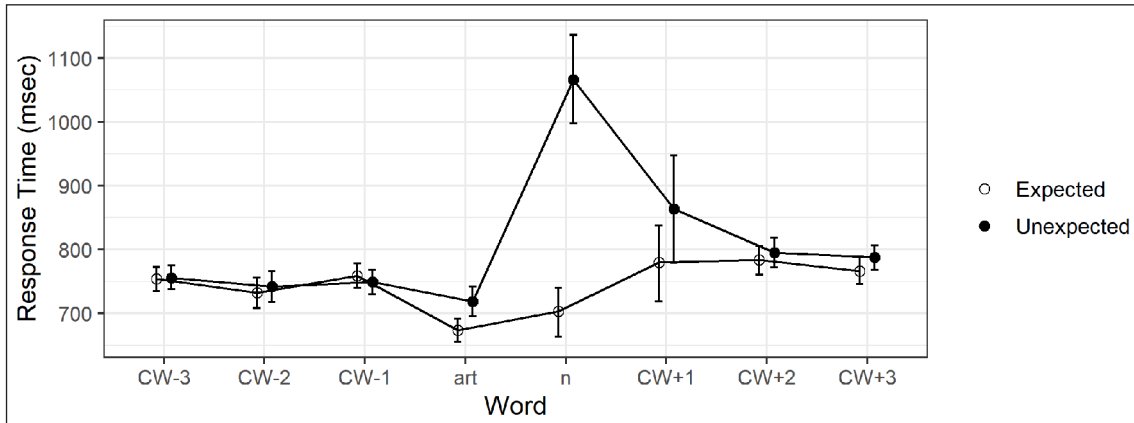
*Error rates.* **Table 2** presents the average error rates for each word by condition. The error rate over all eight regions was low (3.8%). On the target article and noun, error rates were 4.2% and 3.0%, respectively. Two participants whose error rates were two standard deviations above the group average (25.7% and 18.3%) were removed from further analysis. All other participants had average error rates below 10%, with an overall error rate average of 2.8% (article: 3.4%; noun: 2.1%).

**Table 2:** Average error rates across Expectation by word. Standard errors by subject are given in parentheses.

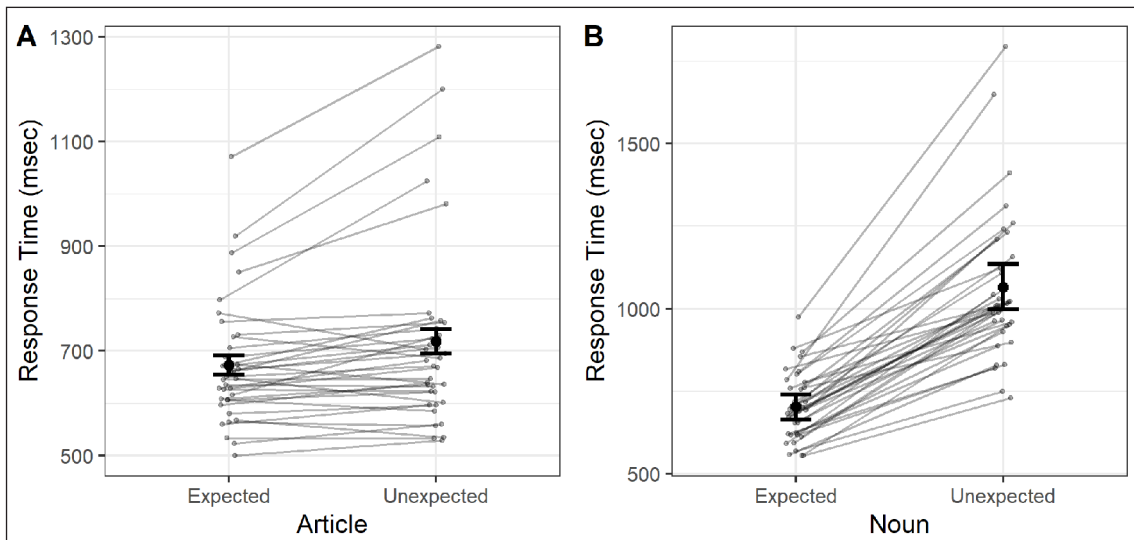
	CW-3	CW-2	CW-1	art	n	CW+1	CW+2	CW+3
<b>Expected</b>	2.6 (2.7)	3.1 (2.9)	3.2 (2.9)	4.0 (3.3)	1.6 (2.1)	3.0 (2.9)	2.9 (2.8)	2.3 (2.5)
<b>Unexpected</b>	3.1 (2.9)	3.1 (2.9)	2.5 (2.6)	2.7 (2.7)	2.5 (2.6)	2.5 (2.6)	3.9 (3.2)	2.0 (2.3)

*Response times.* Error and post-error responses were removed. **Figure 1** shows results by Expectation. RTs for the unexpected condition were significantly slower than the expected condition on not just the noun (Est. = 362.38,  $t = 8.76$ ,  $p < .001$ ) but also on the preceding

article (Est. = 45.51,  $t = 2.91$ ,  $p = .005$ ). Similar results were obtained using noun cloze (article: Est. = -22.92,  $t = -2.94$ ,  $p = .005$ ; noun: Est. = -178.21,  $t = -9.15$ ,  $p < .001$ ). No significant differences were found on regions prior to the article.



**Figure 1:** Response times by region. Error bars indicate difference-adjusted 95% mixed-effect-model-based intervals (Politzer-Ahles, 2017) which can generally be interpreted as indicating that two conditions are likely (but not guaranteed) to be significantly different in a mixed effect model when one condition's interval does not include the other's mean.

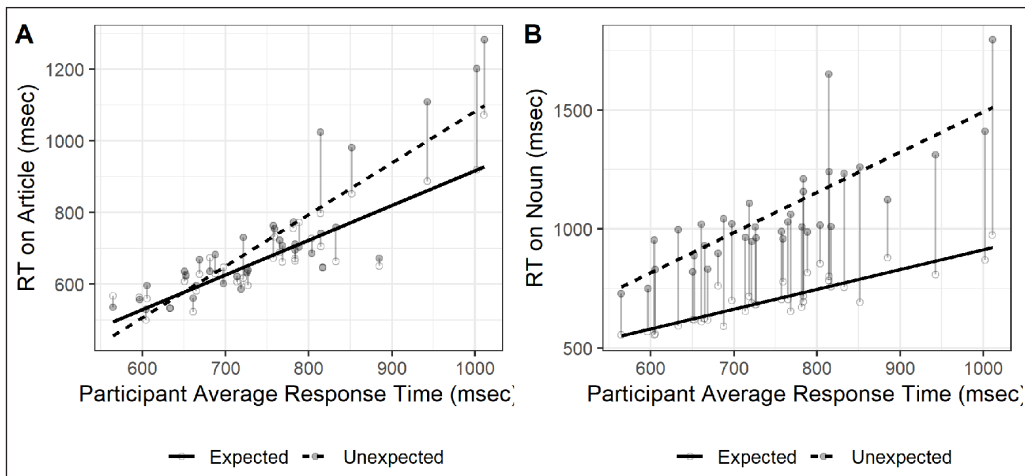


**Figure 2:** Response times on the article (A) and the noun (B) by expectation condition. Individual participant average reading times are shown in addition to the model's conditional mean RTs and difference-adjusted 95% mixed-effect-model-based intervals.

An examination of reading times by participant on the article and noun, shown in **Figure 2**, suggested that the Expectation effect was greater for slower responders. To investigate this, each participant's average response time was calculated based on the average RTs of the three words prior to the article (CW-3, CW-2, and CW-1) over all trials. This predictor was centered (average Participant Average RT: 748 msec) and added to the models above. Significant interactions between Participant Average RT and Expectation were found on both the article ( $t = 2.96$ ,  $p = .004$ ) and the noun ( $t = 4.18$ ,  $p < .001$ ), summarized in **Table 3** and **Figure 3**. Similar significant interactions were found with noun cloze (article:  $t = -2.87$ ,  $p = .006$ ; noun:  $t = -3.00$ ,  $p = .007$ ).

**Table 3:** Summary of four linear mixed effects models for article and noun reading times given Expectation or Noun Cloze crossed with Participant Average RT. P-values are estimated using the Satterthwaite approximation.

		Est	Std.Err	<i>t</i>	<i>p</i>	
<b>Article</b>	Intercept	696.59	11.66	59.72	<.001	***
	Expectation	46.74	13.13	3.56	.001	**
	Participant Average RT	1.20	0.12	10.28	<.001	***
	Expectation:Participant Average RT	0.47	0.16	2.96	.004	**
	Intercept	697.92	14.80	47.15	<.001	***
	Noun Cloze	-21.07	6.77	-3.11	.003	**
	Participant Average RT	1.23	0.15	8.14	<.001	***
	Noun Cloze:Participant Average RT	-0.23	0.08	-2.87	.006	**
<b>Noun</b>	Intercept	885.11	17.71	49.98	<.001	***
	Expectation	362.30	48.56	7.46	<.001	***
	Participant Average RT	1.26	0.16	8.10	<.001	***
	Expectation:Participant Average RT	0.86	0.20	4.18	<.001	***
	Intercept	884.03	14.78	59.82	<.001	***
	Noun Cloze	-179.46	17.38	-10.32	<.001	***
	Participant Average RT	1.26	0.11	12.03	<.001	***
	Noun Cloze:Participant Average RT	-0.43	0.14	-3.00	.007	**



**Figure 3:** Response times on the article (A) and the noun (B) given participant average response time. Individual participant's expected and unexpected averages are plotted and connected by a line to ease identification.

**Figure 2** also suggests that five participants responded very slowly to articles compared to other participants, with an average article RT over 900 msec ( $>1.38$  sd), while also showing large Expectation effects. To investigate whether these slowest responders are the main drivers of expectation and/or its interaction with Participant Average RT on article RTs, an additional model was fit to the data with these five participants removed, leaving 31 participants. Results continue to find a significant effect of Expectation (Est. = 19.28,  $t = 2.09$ ,  $p = .040$ ) and Participant Average RT (Est. = 0.67,  $t = 7.05$ ,  $p < .001$ ), but their interaction is no longer significant (Est. = 0.11,  $t = 0.99$ ,  $p = .323$ ). Additionally, **Figure 3** shows that three participants were responding much slower overall, with Participant Average RTs greater than 900 msec ( $>1.42$  sd), and appear to have an outsized effect on Expectation. To further investigate whether Expectation effects are primarily being driven by these slowest participants, another model was fit to the data excluding these three participants, leaving 33 participants. Again, results find a significant effect of Expectation (Est. = 28.81,  $t = 2.59$ ,  $p = .012$ ) and Participant Average RT (Est. = 0.87,  $t = 6.64$ ,  $p < .001$ ), but their interaction is no longer significant (Est. = 0.24,  $t = 1.96$ ,  $p = .055$ ). Together, these additional analyses suggest that, although an effect of Expectation persists when removing these slow responders, the Expectation effects are magnified by them.

## Discussion

By providing a focal reading time measure, the maze task was able to reveal effects of expectation both on target nouns and the article *a/an* contrast preceding the nouns. Unexpected nouns and their preceding articles, which mismatched the expected noun's required article form, were



responded to more slowly than expected nouns and their preceding articles. Response times on articles and nouns were also found to be inversely related to noun cloze probabilities, with response times decreasing as cloze probabilities increased. This suggests that pre-activation of word form is a more graded effect (DeLong, Urbach, & Kutas, 2005).

Interesting, early predictive effects of article form were magnified for slower responders and smaller with faster responders. This suggests that probabilistic pre-activation of the expected word to the level of phonological form that is required to compute expectations for preceding article form may take time to emerge, consistent with prediction-as-production theories (Pickering & Gambi, 2018). The maze task may be sensitive to these effects because its response times are generally longer than self-paced reading's reading times, eyetracking's fixation durations, or ERP's typical SOAs. This additional time may have aided form prediction. Ito, et al., (2016), for example, found that N400-reduction to form-related words emerged at 700 msec SOA but not 500 msec SOA. A similar effect may have emerged in a more natural and graded fashion here through the variation of individual's average response times. Slower responders may have given themselves more time to reach a form prediction for the upcoming noun and compute its consequences for the preceding article's form, while faster responders were less likely to make these form predictions and consequent computations.

These results demonstrate that the maze task can be sensitive to the predictive use of phonotactic constraints between an expected word and its preceding word, and may be a useful alternative methodology for investigating predictive comprehension mechanisms. A-maze eases the burden of distractor generation and is effective with online crowdsourced participants (Boyce, Futrell, & Levy, 2020). It is hoped that further investigation into predictive mechanisms will be spurred on with this demonstration of the maze task's effectiveness.

---

## Appendix

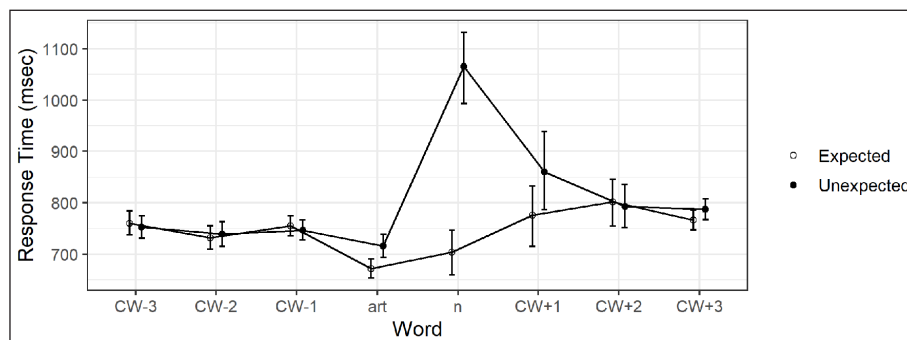
The following reports seven supplemental analyses: 1) RTs including both accurate and inaccurate responses, 2) RTs to nouns given prior article RTs by expectation, 3) expectation given article word form *a/an*, 4) expectation given word sentence position, 5) expectation given article's alternative word, 6) expectation given article's alternative word length, and 7) expectation given participant age.

### A1. Analysis including both accurate and inaccurate responses

All response time analyses reported in the main paper exclude trials where participants initially chose the distractor word. To investigate whether excluding these trials significantly affected response times, a supplemental analysis was conducted including all data, revealing very similar effects to those reported in the main analysis.

**Table A1:** Summary of linear mixed effects models for article, noun, and CW + 1 response times including both correct and incorrect responses. P-values are estimated using the Satterthwaite approximation.

		Est	Std.Err	t	p	
Article	Intercept	694.25	25.00	27.77	< .001	***
	Expectation	44.19	15.25	2.90	.005	**
Noun	Intercept	884.69	29.49	30.00	< .001	***
	Expectation	360.98	42.42	8.51	< .001	***
CW + 1	Intercept	818.24	33.49	24.43	< .001	***
	Expectation	83.36	45.11	1.85	.069	.



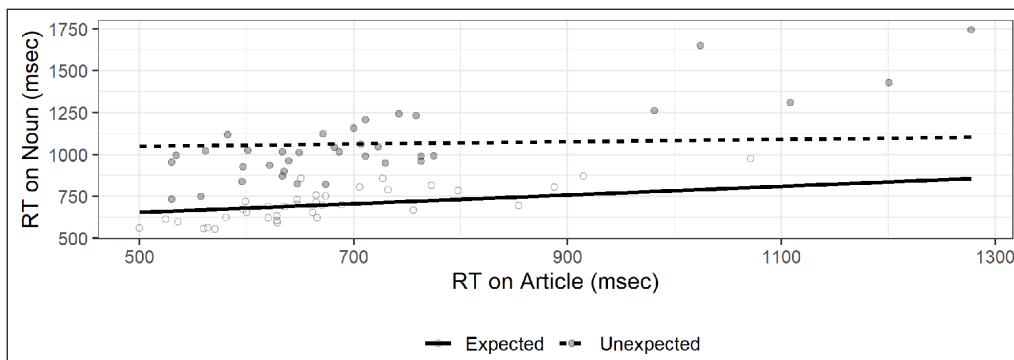
**Figure A1:** Response times including both correct and incorrect responses by region. Error bars indicate difference-adjusted 95% mixed-effect-model-based intervals (Politzer-Ahles, 2017) which can generally be interpreted as indicating that two conditions are likely (but not guaranteed) to be significantly different in a mixed effect model when one condition's interval does not include the other's mean.

## A2. Response times to expected and unexpected nouns given prior article RT

The slowdown in response times by participants on the article may reflect not only an early signal to prediction failure, but also reflect recovery processes. Under this idea, increased response times on unexpected articles should decrease the response time to an unexpected noun. To investigate this possibility, we fit a model to the RTs on the Noun given the RTs on the Article and whether the noun was expected or unexpected. Article RTs were centered prior to model fit. A significant interaction between Article RT and Expectation was discovered, such that expected nouns were read slower the slower comprehenders read their prior article (0.26 msec/Article msec [0.17, 0.35]). RTs on Unexpected nouns, however, showed much less slowdown given the RT on their prior article (0.07 msec/Article msec [0.01, 0.13]), suggesting that the additional time spent on the unexpected article may have lessened the amount of time needed on the unexpected noun.

**Table A2:** Summary of linear mixed effects model for noun response times given their prior article response time. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	<i>t</i>	<i>p</i>	
Intercept	884.13	26.11	33.86	< .001	***
Article RT	355.96	41.85	8.51	< .001	***
Expectation	0.16	0.03	5.95	< .001	***
Article RT:Expectation	-0.19	0.05	-3.62	< .001	***



**Figure A2:** Response times to nouns given response times to their prior article. Individual participant's expected (light unfilled) and unexpected (dark filled) averages are also shown.

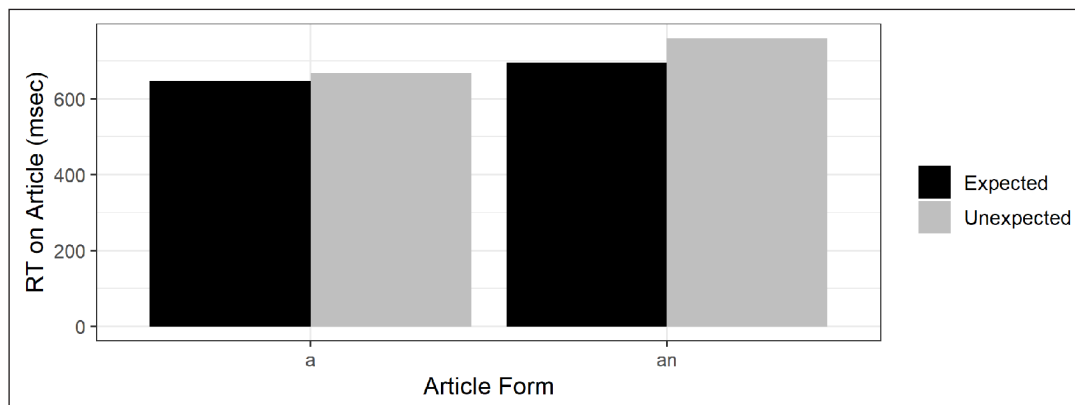
## A3. Expectation by Word Form (a/an)

Because the English indefinite article forms differ in length and frequency, we might expect differences in their effect on expectation. This was investigated in a model including article form. Article Form and Expectation affected RTs, with a marginal interaction. Examining Expectation

within Article Form on article RTs found a significant effect of Expectation on *an* forms (Est. = 64.9,  $t = 3.29$ ,  $p = .002$ ) but not *a* forms (Est. = 22.1,  $t = 1.04$ ,  $p = .301$ ).

**Table A3:** Summary of linear mixed effects model for article response times given article form. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	$t$	$p$	
Intercept	692.46	25.02	27.67	< .001	***
Expectation	43.50	15.44	2.82	.006	**
Article Form	70.13	29.30	2.39	.018	*
Expectation:Article Form	42.87	26.34	1.628	.107	



**Figure A3:** Average response times to expected and unexpected articles given their article form, *a* or *an*.

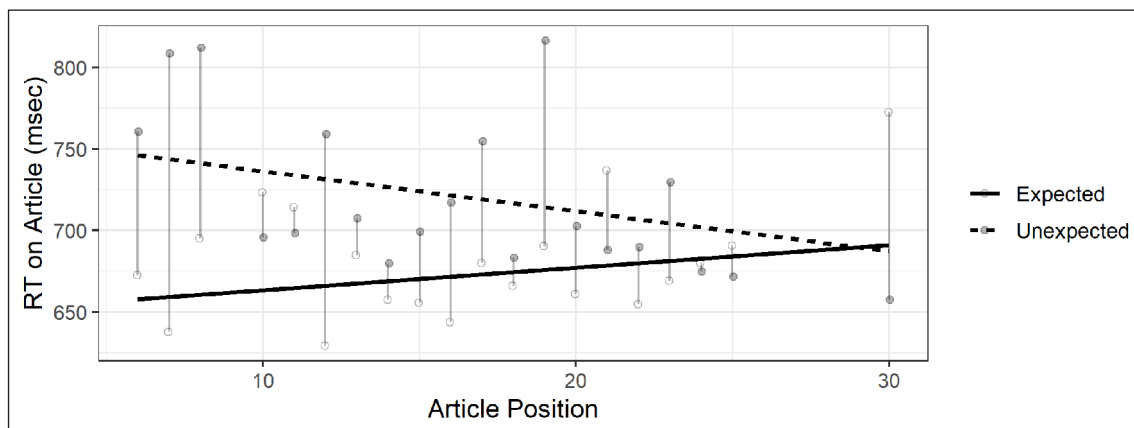
Article Form was also found to marginally affect noun RTs (Est. = 89.11,  $t = 1.72$ ,  $p = .089$ ), though the interaction between Expectation and Article Form was not significant (Est. = 61.91,  $t = 0.84$ ,  $p = .405$ ).

#### A4. Expectation by Word Position

It is well known that comprehenders tend to make faster responses as they read. To examine whether the position of the article in the sentence affected expectation on article RTs, article position (centered) was included in a model. Article Position, however, did not show a significant effect on article RTs nor was there a significant interaction.

**Table A4:** Summary of linear mixed effects model for article response times given their linear position in the sentence. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	t	p	
Intercept	695.51	24.81	28.04	< .001	***
Expectation	41.98	15.74	2.67	.010	*
Article Position	-0.54	1.27	-0.42	.674	
Expectation:Article Position	-3.82	2.48	-1.54	.127	



**Figure A4:** Response times to articles by their sentence position. Average RTs given article sentence position for expected and unexpected articles are also shown.

A similar analysis with noun position on noun RTs found a significant effect of Expectation (Est. = 363.47,  $t = 8.78$ ,  $p < .001$ ), but did not find a significant effect of position (Est. = 2.78,  $t = 0.84$ ,  $p = .401$ ) or a significant interaction (Est. = 6.53,  $t = 0.90$ ,  $p = .373$ ).

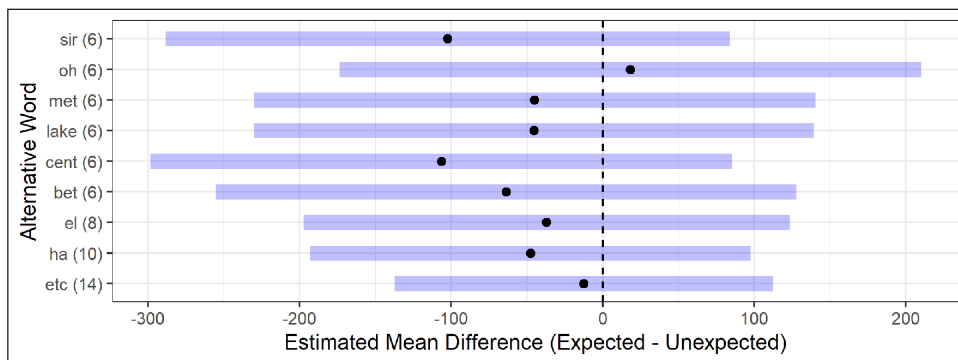
## A5. Expectation by Alternative Word to Article

The maze task requires comprehenders to choose between two alternative words to continue reading and therefore properties of the alternative word form could affect response times. Because alternatives are matched as closely as possible in terms of word length and frequency, A-maze may have had difficulty selecting appropriate alternatives for indefinite articles. 45 different alternative words were used in this study, with some being repeated more often than others (minimum repetition: 2, maximum repetition: 14). The distribution of alternative repetitions is given in Table S5.

**Table A5:** Count of article alternative words given their repetition in the study.

Number of repetitions	2	4	6	8	10	14
Number of alternative words	27	9	6	1	1	1

A linear mixed effects model of article RTs including expectation and alternative form was fit. The model found a main effect of Expectation ( $F(1) = 6.97$ ,  $p = .011$ ) but no significant effect of Alternative ( $F(44) = 1.35$ ,  $p = .141$ ) and no significant interaction ( $F(44) = 0.81$ ,  $p = .751$ ).



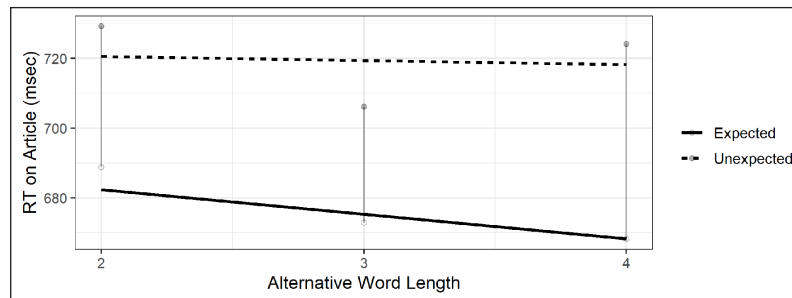
**Figure A5:** Estimated contrast of expected minus unexpected article RTs for the nine most repeated alternative forms to articles. Number of alternative word repetitions is given in parentheses.

## A6. Expectation by Alternative Word to Article Length

Because there was a large number of different alternative forms, we also investigated whether a simple measure of alternative length could be shown to affect article RTs. Alternative Length was included in a model with Expectation. Neither Alternative Length nor an interaction was found to be significant.

**Table A6:** Summary of linear mixed effects model for article response times given the length of the alternative. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	t	p	
Intercept	696.22	24.82	28.05	< .001	***
Expectation	45.56	14.25	3.20	.003	**
Alternative Length	-4.08	7.43	-0.55	.585	
Expectation:Alternative Length	5.90	12.08	0.49	.625	



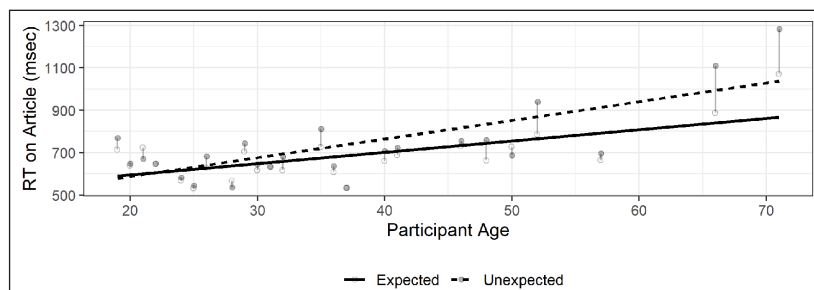
**Figure A6:** Response times to articles by their alternative word's length. Average RTs given alternative word length for expected and unexpected articles are also plotted.

### A7. Expectation by Participant Age

Because older participants tend to have slower response times, we might expect differences in participant age to show an effect on expectation. This was investigated in a model including each participant's age. Age was centered prior to analysis. Both Expectation and Age affected article RTs and there was a significant interaction, with older participants showing a larger difference between expected and unexpected articles compared to younger participants. Table A7b below provides additional detail, including both participant age and their average response times and is ordered by the average RT different between unexpected and expected articles for each participant.

**Table A7a:** Summary of linear mixed effects model for article response times given a participant's age. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	t	p	
Intercept	696.64	19.10	36.47	<.001	***
Expectation	45.63	13.64	3.35	<.001	***
Age	7.07	1.40	5.06	<.001	***
Expectation:Age	3.44	0.89	3.88	<.001	***



**Figure A7a:** Response times to articles by participant age. Average RTs given participant age for expected and unexpected articles are also shown.

**Table A7b:** Participant age, average response time, expected article average RTs, unexpected article average RTs, unexpected article average RTs, and the difference between unexpected and expected article RTs. The table is ordered by difference, from highest to lowest.

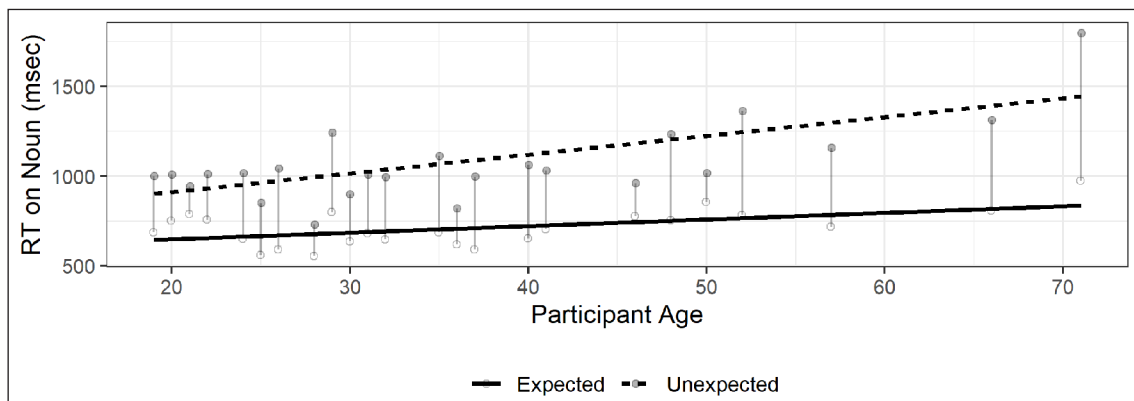
Participant	Age	Part.Avg RT	Expected	Unexpected	Difference	Participant	Age	Part.Avg RT	Expected	Unexpected	Difference
<b>1</b>	52	1002	919	1201	281	<b>19</b>	25	604	500	529	29
<b>2</b>	52	814	798	1024	227	<b>20</b>	36	651	608	635	27
<b>3</b>	66	943	888	1109	221	<b>21</b>	46	759	731	754	23
<b>4</b>	71	1011	1071	1282	211	<b>22</b>	20	885	651	672	21
<b>5</b>	35	852	851	981	130	<b>23</b>	19	782	756	773	17
<b>6</b>	32	722	616	730	113	<b>24</b>	24	665	580	596	16
<b>7</b>	48	833	663	758	96	<b>25</b>	30	714	607	622	15
<b>8</b>	19	758	673	763	91	<b>26</b>	22	817	645	647	1
<b>9</b>	26	688	629	682	53	<b>27</b>	37	633	534	533	-1
<b>10</b>	40	768	662	707	45	<b>28</b>	31	726	635	632	-3
<b>11</b>	35	727	597	639	43	<b>29</b>	20	652	627	622	-5
<b>12</b>	32	783	671	711	40	<b>30</b>	25	596	563	557	-6
<b>13</b>	30	669	628	668	40	<b>31</b>	24	718	607	586	-21
<b>14</b>	24	661	523	560	37	<b>32</b>	28	565	568	535	-33
<b>15</b>	29	815	706	742	37	<b>33</b>	21	681	674	636	-38
<b>16</b>	32	605	560	597	37	<b>34</b>	50	804	727	686	-41
<b>17</b>	41	765	689	723	34	<b>35</b>	52	697	647	601	-46
<b>18</b>	57	783	665	695	30	<b>36</b>	21	789	773	703	-69



Age also showed a significant effect on noun RTs, along with Expectation and their interaction. Older participants again showed a larger difference between expected and unexpected articles compared to younger participants.

**Table A7c:** Summary of linear mixed effects model for noun response times given a participant's age. P-values are estimated using the Satterthwaite approximation.

	Est	Std.Err	t	p	
Intercept	885.06	23.95	36.95	< .001	***
Expectation	362.62	38.60	9.40	< .001	***
Age	7.07	1.47	4.81	< .001	***
Expectation:Age	6.80	1.76	3.85	< .001	***



**Figure A7b:** Response times to nouns by participant age. Average RTs given participant age for expected and unexpected articles are also shown.

## Acknowledgements

This research was supported by the Fellow's Research Fund at St. Hugh's College.

## Competing Interests

The author has no competing interests to declare.

## References

- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, *17*(3), 364–390. DOI: [https://doi.org/10.1016/0010-0285\(85\)90013-1](https://doi.org/10.1016/0010-0285(85)90013-1)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze Made Easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082. DOI: <https://doi.org/10.1016/j.jml.2019.104082>
- Cutter, M. G., Martin, A. E., & Sturt, P. (2020). Readers detect an low-level phonological violation between two parafoveal words. *Cognition*, *204*: 104395. DOI: <https://doi.org/10.1016/j.cognition.2020.104395>
- 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. DOI: <https://doi.org/10.1038/nm1504>
- Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, *80*(1), 27–38. DOI: <https://doi.org/10.1093/biomet/80.1.27>
- Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163–171. DOI: <https://doi.org/10.3758/BRM.41.1.163>
- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(5), 1461–1469. DOI: <https://doi.org/10.1037/a0036756>
- Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *Proceedings of NAACL*. DOI: <https://doi.org/10.18653/v1/N18-1108>
- Heinze, G., Ploner, M., & Jiricka, L. (2020). logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.24. <https://CRAN.R-project.org/package=logistf>
- 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. DOI: <https://doi.org/10.1016/j.jml.2015.10.007>
- Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*. 107291. DOI: <https://doi.org/10.1016/j.neuropsychologia.2019.107291>
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). Why the A/AN prediction effect may be hard to replicate: a rebuttal to DeLong, Urbach, and Kutas (2017). *Language, Cognition and Neuroscience*, *32*(8), 974–983. DOI: <https://doi.org/10.1080/23273798.2017.1323112>
- Kuznetsova, A., Brockhoff, P. B., & Christensen R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: <https://doi.org/10.18637/jss.v082.i13>

- Lenth, Russell V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.2-1. <https://CRAN.R-project.org/package=emmeans>
- Martin, C. D., Branzi, F.M., & Bar, M. (2018). Prediction is Production: the missing link between language production and comprehension. *Scientific Reports* 8(1), 1079. DOI: <https://doi.org/10.1038/s41598-018-19499-4>
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras, & M. A. Just (Eds.). *New methods in reading comprehension*. Hillsdale, NJ: Earlbaum.
- Nieuwland, M. S., Arkipova, Y., & Rodríguez-Gómez, P. (2020). Anticipating words during spoken discourse comprehension: A large-scale, pre-registered replication study using brain potentials. *Cortex*, 133, 1–36. DOI: <https://doi.org/10.1016/j.cortex.2020.09.007>
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthum, S., Bartolozzi, F., Kogan, V., Ito, A., Meziere, D., Barr, D., Rousselet, G., Ferguson, H., Busch-Moreno, S., Fu, X., Kulakova, E., Tuomainen, J., Husband, E. M., Donaldson, D., Kohút, Z., Rueschemeyer, S.-A., & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, 7, e33468. DOI: <https://doi.org/10.7554/eLife.33468.024>
- 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: <https://doi.org/10.1016/j.brainres.2009.07.042>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. DOI: <https://doi.org/10.1016/j.jesp.2017.01.006>
- Pickering, M. J. & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002. DOI: <https://doi.org/10.1037/bul0000158>
- Politzer-Ahles, S. (2017). An extension of within-subject confidence intervals to models with crossed random effects. *The Quantitative Methods for Psychology*, 13(1), 75–94. DOI: <https://doi.org/10.20982/tqmp.13.1.p075>
- Van Berkum, J. J., 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. DOI: <https://doi.org/10.1037/0278-7393.31.3.443>
- 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. DOI: <https://doi.org/10.1162/0898929041920487>
- Witzel, J. & Forster, K. (2014). Lexical co-occurrence and ambiguity resolution. *Language, Cognition and Neuroscience*, 29(2), 158–185. DOI: <https://doi.org/10.1080/01690965.2012.748925>
- Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, 41(2), 105–128. DOI: <https://doi.org/10.1007/s10936-011-9179-x>

