Word predictability and frequency effects in a rational model of reading

Klinton Bicknell[1] (kbicknell@ucsd.edu) & Roger Levy[2] (rlevy@ucsd.edu)
[1]Department of Psychology, UC San Diego, La Jolla, CA, USA
[2]Department of Linguistics, UC San Diego, La Jolla, CA, USA

Abstract

This paper presents results from the first rational model of eye movement control in reading to make predictions for the full range of the eye movement record. The model identifies the text through Bayesian inference and makes eye movement decisions to maximize the efficiency of text identification, going beyond leading approaches which select model parameters to maximize the fit to human data. Two simulations with the model demonstrate that it can produce effects of word predictability and frequency on eye movements in reading similar to those produced by humans, providing evidence that many properties of human reading behavior may be understood as following from the nature of efficient text identification.

Keywords: eye movements; reading; rational analysis; computational modeling

Introduction

During reading, comprehenders must decide when and where to move their eyes 3–4 times every second. Over the past decades, it has been demonstrated that comprehenders make these rapid, fine-grained decisions by combining information from a range of sources including visual input, the motor system, and linguistic knowledge (for reviews see Rayner, 1998, 2009), making reading one of the most complex learned tasks that humans face every day. Gaining a better understanding of this process promises to yield insights about how readers deploy linguistic knowledge for real-time comprehension as well as about how humans learn to perform complex tasks more generally. In this paper, we present the first rational model of eye movement control in reading that makes predictions for the full range of the eye movement record. We model readers as performing Bayesian inference on the identity of the text, combining their probabilistic language knowledge (the prior) with noisy perceptual input about the text (the likelihood) to form and repeatedly update a posterior distribution over the possible text identities. The model uses a parameterized behavior policy for determining when and where to move the eyes, which is sensitive to the posterior distribution over the text, and with parameters selected to optimize identification efficiency. We evaluate the model by examining the effects it produces for two linguistic variables: word frequency and predictability. We present the results of two simulations showing that the model produces effects of these variables similar to those of humans, across four different eye movement measures reflecting both the locations and durations of fixations. The success of the model in deriving these effects from principles of probabilistic inference and rational action suggests that many aspects of human reading behavior may be profitably understood as properties of the set of efficient solutions to the problem of reading.

This model goes beyond leading models of eye movement control in reading such as E-Z Reader (Reichle, Rayner, & Pollatsek, 2003) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) in two ways. First, while those models select parameters to maximize the fit to human data, the current work selects parameters to optimize the efficiency of reading, here characterized as rapid and accurate identification of the contents of the text. To the extent that the model behavior reproduces effects seen in human data, then, it enables understanding those effects as resulting from the properties of efficient solutions to the task. Second, the current work includes a model of the process of identification from visual input, and in so doing derives effects of linguistic variables (such as word frequency and predictability) as resulting from efficient identification, while models such as E-Z Reader and SWIFT directly specify the effects of linguistic variables on eye movement behavior through functions whose form is stipulated exogenously to the model. Modeling identification from visual input should allow for the model to be used to understand a range of effects that are known to influence eye movements but which leading approaches cannot capture, such as information density within words (Hyönä, Niemi, & Underwood, 1989), word misidentification (Slattery, 2009; Levy, Bicknell, Slattery, & Rayner, 2009), and visual neighborhoods (Pollatsek, Perea, & Binder, 1999). The model also goes beyond the only previous rational model of eye movement control in reading, Mr. Chips (Legge, Hooven, Klitz, Mansfield, & Tjan, 2002), in making predictions about not only the location of fixations but also their duration, which is important to gaining a full understanding of a range of effects on eye movements in reading, especially the effects of linguistic variables.

In the following section, we describe our rational framework for reading and the details of our model of eye movement control in reading. We then focus the remainder of the paper on using the model to understand the effects on eye movements in reading of word frequency and predictability, two of the most reliable linguistic effects in the eye movement record. We first use the model qualitatively to provide explanations for why the effects of these variables seen empirically should result from efficient reading behavior, and then present the quantitative results of two simulations demonstrating that these effects are evident in the model's behavior.

Reading as Bayesian inference

In the proposed framework, we model the goal of reading as efficient text identification. While it is clear that this is not all that readers do – inferring the underlying structural relationships among words in a sentence and discourse relationships between sentences that determine text meaning is a fundamental part of most reading – all reader goals necessarily involve identification of at least part of the text, so we take text identification to be a reasonable first approximation. There are two sources of information relevant to this goal: visual input and language knowledge, which the model combines via Bayesian inference. Specifically, it begins with a prior distribution over possible identities of the text given by its language model, and combines this with noisy visual input about the text at the eyes' position (giving the likelihood term) to form a posterior distribution over the identity of the text taking into account both the language model and the visual input obtained thus far. On the basis of the posterior distribution, the model then decides whether or not to move its eyes (and if so where to move them to) and the cycle repeats.

An implemented model in this framework must formalize a number of pieces of the reading problem, including the possible actions available to the reader and their consequences, the nature of visual input, the nature of language knowledge, a means of combining visual input with prior expectations about the form and structure of the text, and a behavior policy determining how the model will choose actions on the basis of its posterior distribution over the identity of the text. In the remainder of this section, we present the details of our formalizations of these pieces.[1]

Formal problem of reading: Actions

We assume that on each of a series of discrete timesteps, the model obtains visual input around the current location of the eyes, and then chooses between three actions: (a) continuing to fixate the currently fixated position, (b) initiating a saccade to a new position, or (c) stopping reading. If the model chooses option (a), time simply advances, and if it chooses option (c), then reading immediately ends. If a saccade is initiated (b), there is a lag of two timesteps, representing time required to plan a saccade, during which the model again obtains visual input around the current position, and then the eyes move toward the intended target. Because of motor error, the actual landing position of the eyes is normally distributed around the intended target with standard deviation given by a linear function of the intended distance, with parameters taken from Engbert et al. (2005).[2]

Noisy visual input

The visual input obtained by a reader on a given timestep is generated from the following process, independently for each character position. Each letter is represented as a 26-dimensional vector, where a single element is 1 and the others are zeros, and visual input about a letter is a sample from a 26-dimensional Gaussian with a mean equal to the letter's true identity and a diagonal covariance matrix $\Sigma = \lambda^{-1}I$, where λ is the reader's visual acuity at that position. Higher visual acuity, then, means a lower sample variance, yielding higher quality visual input. We use the visual acuity function from Engbert et al. (2005), in which λ decreases exponentially with retinal eccentricity and decreases asymmetrically, falling off more slowly to the right than the left.[3] In order to scale the quality of visual information, we multiply each acuity λ by the overall visual input quality Λ (values given in the simulations below.) Visual input about non-alphabetic characters is veridical knowledge of their identity. Visual input is limited to the 19 character positions with the highest acuity (eccentricities between -7 and 12), roughly corresponding to estimates that readers of English obtain useful information from about 19 characters, and more from the right of fixation than the left (Rayner, 1998). Note that in the model each letter is equally confusable with all others, following Norris (2006, 2009), but ignoring work on letter confusability (which could be added to future model revisions; Engel, Dougherty, & Brian Jones, 1973; Geyer, 1977).

Language knowledge

In general, any generative model of linguistic knowledge that assigns probabilities to text can be used as the prior distribution on the identity of the text. For the simulations in this paper, we use very simple probabilistic models of language knowledge: word *n*-gram models (Jurafsky & Martin, 2009). These models encode the probability of each word conditional on the $n-1$ previous words. While this is obviously a crude representation of the rich knowledge of language that human readers have, it serves here to illustrate the qualitative effects of using linguistic context in reading.

Inference about text identity

Given both visual input and language knowledge, the model makes inferences about the identity of the text w via standard Bayesian inference, where the prior is given by the probability of generating text identity w from the language model and the likelihood is the probability of generating the visual input $\mathcal{I}$ from text with identity w under the visual input model:

$$p(w|\mathcal{I}) \propto p(w)p(\mathcal{I}|w).$$

Behavior policy

The model uses a simple policy with two parameters, α and β, to decide between actions based on the marginal probability m of the most likely character c in each position j,

$$m(j) = \max_c p(w_j = c)$$

where w_j indicates the character in the jth position. A high value of m indicates relative confidence about the character's identity, and a low value relative uncertainty.

Given the values of this statistic m, the model decides between four possible actions, as illustrated in Figure 1. If the

[1] See Bicknell and Levy (2010b) for further computational details.

[2] In the terminology of the literature, the model has only random motor error (variance), not systematic error (bias). Following Engbert and Krügel (2010), systematic error may arise from Bayesian estimation of the best saccade distance.

[3] While we call refer to it here as visual acuity, it is clear from the asymmetric nature of this function that it also has an attentional component. For now, however, we make the simplifying assumption that it is unchanging over time.

(a)	$m = [.6, .7, \mathbf{.6}, .4, .3, .6]$:	Keep fixating (3)
(b)	$m = [.6, .4, \mathbf{.9}, .4, .3, .6]$:	Move back (to 2)
(c)	$m = [.6, .7, \mathbf{.9}, .4, .3, .6]$:	Move forward (to 6)
(d)	$m = [.6, .7, \mathbf{.9}, .8, .7, .7]$:	Stop reading

Figure 1: Values of m for a 6 character text under which a model fixating position 3 would take each of its four actions, if $\alpha = .7$ and $\beta = .5$.

value of this statistic for the current position of the eyes is less than the parameter α, the model chooses to continue fixating the current position (1a). Otherwise, if the value of $m(j)$ is less than the parameter β for some leftward position, the model initiates a saccade to the closest such position (1b). If no such positions exist to the left, then the model initiates a saccade to n characters past the closest position to the right for which $m(j) < \alpha$ (1c).[4] Finally, if no such positions exist to the right, the model stops reading (1d). Intuitively, then, the model reads by making a rightward sweep to bring its confidence in each character up to α, but pauses to move left to reread any character whose confidence falls below β.

Predictability and frequency in rational reading

The general findings about the effects of word predictability and frequency on eye movements in reading can be summarized relatively simply: words that are less predictable and lower frequency tend to receive more and longer fixations (Rayner, 1998, 2009). Here we describe intuitions for why our model should qualitatively reproduce these effects.

Predictability

The basic intuition for why the model should produce effects of word predictability is very closely related to the reason for frequency effects in isolated word recognition reaction times given by Norris (2006, 2009). In short, the lower the prior probability of a word, the more visual input about it is needed to become confident in its identity. A bit more formally, this intuition is clearest if we make the simplifying assumption that prior to obtaining any visual information about a word, the model has near-veridical knowledge of the preceding context. In that case, the probability of the true identity of the word is given by the word's predictability in context π. Visual input about the word will then (on average) increase the probability of the word's true identity under the model's beliefs. Recall that under our behavior policy, the eyes will remain in this position until the model's confidence in the identity of the character at that position exceeds the threshold α. Because information is being obtained about the entire word simultaneously, the probability of the identity of the fixated character is closely tied to the identity of the entire word. Specifically, the model's confidence in the identity of the word gives a lower bound on the model's confidence in the identity of a character within that word. Thus, the initial probability of the true identity of the fixated character will start at or above the initial probability π of the true word, and – when the word is identified correctly – the model's confidence about the identity of the fixated character is likely to reach the threshold α near the same time that confidence about the identity of the fixated word reaches the threshold. As a consequence, the amount of visual input that is needed to reach the threshold which initiates a saccade is largely a function of the distance between π and α. For more predictable words, π is closer to α, so less visual input will be needed on average to reach α, translating into shorter and fewer fixations on the word.

Frequency

The most obvious intuition for the effect of frequency in the model is parasitic on the effect of predictability: words that are lower frequency are less predictable on average. Thus, as with words of higher predictability, there should be on average shorter and fewer fixations on words of high frequency.

Simulation 1: full model

We now assess the effects of word predictability and frequency that the model does in fact produce. We use the model to simulate reading of a modified version of the Schilling corpus (Schilling, Rayner, & Chumbley, 1998) of typical sentences used in reading experiments. The arguments just described predict qualitatively that the model will make more and longer fixations on words of lower predictability and frequency. In addition, we quantitatively compare the model's frequency effects to those of human readers of the Schilling corpus, which have been reported by Pollatsek et al. (2006).

Methods

Model implementation We implemented our model with weighted finite-state automata (wFSAs) using the OpenFST library (Allauzen, Riley, Schalkwyk, Skut, & Mohri, 2007). While inference in the wFSA is exact, for efficiency we used Monte Carlo sampling from the wFSA to estimate the model's confidence m in each character position.

Model parameters and language model We set the overall visual input quality Λ to 4. The model's language knowledge was an unsmoothed bigram model created using a vocabulary set consisting of the 500 most frequent words in the British National Corpus (BNC) as well as all the words in our test corpus. From this vocabulary, we counted every bigram in the BNC for which both words were in vocabulary. Due to the intense computation required for exact inference, we then trimmed this set by removing rare bigrams that occur less than 200 times (except that we do not trim any bigrams that occur in our test corpus). This left a set of about 19,000 bigrams, from which we constructed the bigram model.

Optimization of policy parameters We define reading efficiency E to be an interpolation of speed and accuracy

$$E = (1 - \gamma)L - \gamma T$$

[4] The role of n is to ensure that the model does not center its visual field on the first uncertain character. For the present simulations, we did not attempt to optimize this parameter, but fixed n at 3.

where L is the log probability of the true identity of the text under the model's beliefs at the end of reading, T is the number of timesteps before the model stopped reading, and γ gives the relative value of speed. For the present simulations, we use $\gamma = .05$, which produces reasonably accurate reading. To find optimal values of the policy parameters α and β for this definition of efficiency, we use the PEGASUS method (Ng & Jordan, 2000) to transform this stochastic optimization problem into a deterministic one on which we can use standard optimization algorithms. We then use coordinate ascent (in logit space) to find the optimal values of α and β. This procedure resulted in optimal values $\alpha = .88$ and $\beta = .98$.[5]

Test corpus To ensure that results did not depend on smoothing, we tested the model only on sentences from the Schilling corpus in which every bigram occurred in the BNC. Unfortunately, only 8 of the corpus sentences initially met this criterion, so we made single-word changes to 25 more (mostly proper names and rare nouns), producing a total of 33 sentences for which every bigram occurred in the BNC.

Analysis We used the model to perform 50 stochastic simulations of the reading of our modified version of the Schilling corpus. For each run, we calculated four standard eye movement measures for each word in the corpus: first fixation duration, gaze duration (defined to be the sum of all first pass fixations), skipping probability (whether or not word was directly fixated), and refixation probability (the probability of more than one first pass fixation). We then averaged each of these four measures for each word token in the corpus, yielding a single mean value for each measure for each word.

In order to compare the fixation duration measures to humans, we converted the model's timesteps into milliseconds. We performed this scaling by multiplying the duration of each fixation by a conversion factor set to be equal to the mean human gaze duration divided by the mean model gaze duration for the highest frequency bin. That is, we scaled the model predictions to exactly match the human mean for gaze durations in the highest frequency bin.

Results

For each word in our modified version of the Schilling corpus, we defined its predictability to be its probability under the bigram language model, and we defined its frequency to be its overall probability in the data from which the bigram language model was constructed.

Predictability Figure 2 (red lines) shows the effect of predictability on the four aggregate measures. As predicted by both the intuition given above, and in agreement empirical human data, there are shorter fixations, more skipping, and fewer refixations for more predictable words.

[5]It may at first seem puzzling that $\alpha < \beta$. However, this is a general property of optimal behavior for the model. While saccades to leave a character are initiated as soon as confidence $m > \alpha$, because of the saccade execution delay, m is usually substantially higher than α when the eyes leave the character. Hence, it is a reasonable strategy for the threshold for regressions β to be accordingly higher. See also Bicknell and Levy (2010b) for further discussion.

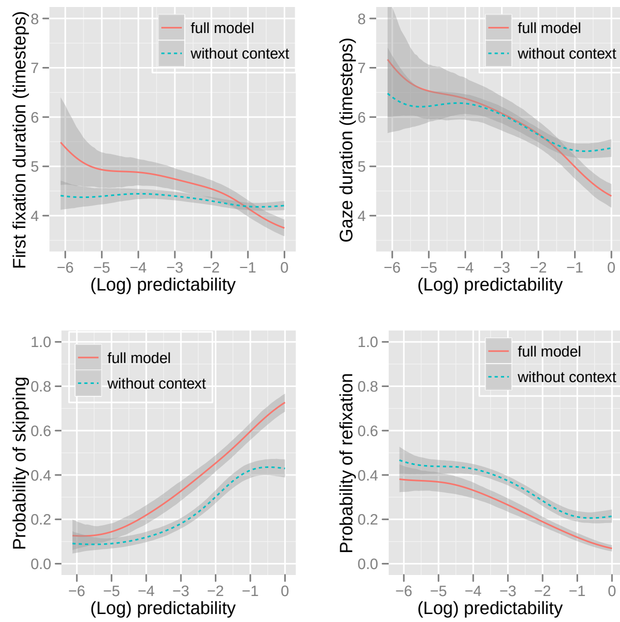

Figure 2: Effects of word predictability in both models on first fixation durations, gaze durations, the rate of skipping, and the rate of making a refixation, as estimated using Gaussian kernel regression with standard deviation equal to 1/8th of the range of log-predictability values. The 95% confidence intervals are bootstrapped from 1000 dataset replicates.

Frequency Figure 3 (red lines) shows the effects of frequency (binned by rounding down, to facilitate comparison to Pollatsek et al., 2006) on the four aggregate measures. The results across all four measures show a reasonable quantitative fit to the human data (blue lines). Further, comparing the overall size of the effect (i.e., the difference of the highest and lowest frequency bins) of the model to the human data shows a striking fit in effect direction and magnitude for all four measures. One unpredicted result here, however, is that the effect of frequency on the duration measures does not appear completely monotonic.

Discussion

In summary, these results demonstrate that effects of predictability and frequency in the model's behavior resemble that of human readers in many respects. Predictability effects on all four aggregate measures are monotonic and in the same direction as predicted. Frequency effects on all four measures are in the same direction as predicted, and the total magnitude of the effect is quite similar to that displayed by human readers, despite the fact that we have not made any attempt to fit the human data, excepting only the scaling parameter that converts model timesteps to milliseconds. Overall quantitative fits on all four measures showed reasonable agreement to human data, but the fixation duration measures displayed

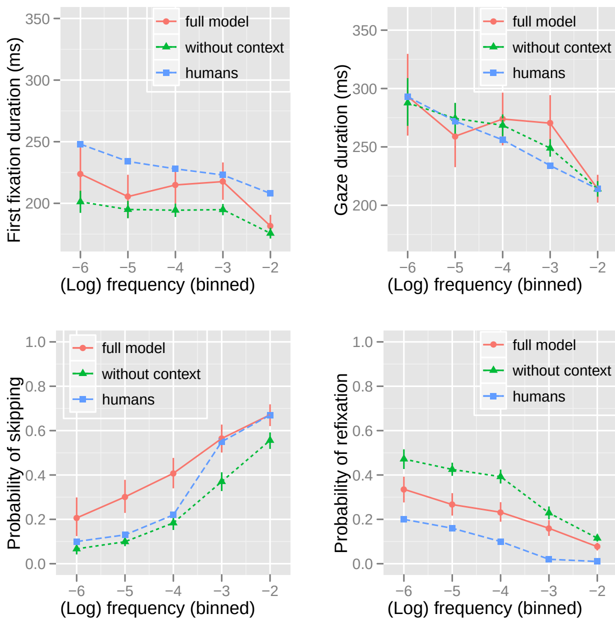

Figure 3: Effects of word frequency in both models on first fixation durations, gaze durations, the rate of skipping, and the rate of making a refixation. The 95% confidence intervals are bootstrapped from 10000 dataset replicates. Mean values from human readers of the Schilling corpus reported by Pollatsek et al. (2006) are shown for comparison.

some non-monotonicity.

It is perhaps unsurprising that predictability seems to have the most consistent effect, given the large role that predictability plays in the model, and the relatively straightforward predictions made previously. More surprising are the apparent non-monotonicities in the predictions for how the fixation duration measures should vary with respect to word frequency. One possibility is that these arise from our artificial removal of many low-frequency words from the language model, which may have meant that some of the low-frequency words in the Schilling corpus had artificially few visual neighbors, yielding an anti-frequency effect. The next simulation investigates this hypothesis.

Simulation 2: Model without context

The main goal of Simulation 2 is to explore the possibility that removing low frequency words from the model's vocabulary (which was necessary for computational efficiency) contributed to the non-monotonicities we observed in the effects of word frequency on fixation durations. Our strategy is to simplify the language model, which makes the computations faster to carry out, allowing for the use of a larger vocabulary. Specifically, we replace the previous bigram language model, which made use of linguistic context, with a unigram language model that includes only word frequency information and cannot make use of linguistic context. This simplified language knowledge also allows us to test how the model's predictions change when it can no longer make use of linguistic context to help recognize words.

Methods

Except the following, the methods were identical to those of Simulation 1. We replaced the bigram language model with a unigram language model. Training was performed in the same manner, except that instead of including only the most common 500 words in the BNC, we included all words that occur at least 200 times (corresponding to a frequency of 2 per million; about 19,000 words). Finally, we increased the overall visual input quality Λ from 4 to 10. Because the new language model gives poorer information about the text, more visual input is needed to reach similar levels of confidence in word identities. Increasing the overall input quality to 10 results in the new model taking a similar number of timesteps to read a sentence as the previous model.

Results and discussion

Predictability Figure 2 (green lines) shows the effect of predictability on the four aggregate measures for the model without context. Because the model does not make use of linguistic context in identifying words, any apparent effects of predictability must reflect effects of other variables correlated with predictability (e.g., frequency and length). We can then use these results as a baseline to determine the amount of the full model's apparent predictability effect that was in fact driven by predictability. The results across all four measures show that predictability effects are smaller for this model without context, indicating that the full model's use of context was important in producing its predictability effects.

Frequency Figure 3 (green lines) shows the effect of frequency on the four aggregate measures. Across all four measures, the size of the frequency effect in this model also shows a reasonable quantitative fit to human data, although the refixation rates and first fixation durations are about twice as far from human data as the full model. As with the full model, however, the direction and magnitude of all frequency effects is a close match to human data. The higher refixation rate and lower word skipping rate of this model relative to the full model likely reflect the model's poorer language knowledge (cf. Bicknell & Levy, 2010a). Finally, and most importantly, we see that the problem of non-monotonicity is substantially reduced for first fixation durations and completely eliminated for gaze durations, supporting our argument that trimming the vocabulary may have been responsible for some of the non-monotonicity in the previous simulation results.

General discussion

In this paper, we presented the first rational model of eye movement control in reading to make predictions for the entirety of the reading record. We gave intuitions for why it should produce effects of word predictability and frequency qualitatively similar to those produced by human readers, and presented two simulations empirically testing the effects of these variables on model behavior. Simulation 1, using a full version of the model with parameters selected to maximize the agent's reading efficiency, demonstrated that the model yields effects of frequency and predictability that are qualitatively – and in frequency's case, quantitatively – similar to those of human readers, though the predictions for fixation durations on words of intermediate frequency did not appear completely monotonic. We hypothesized that these non-monotonicities may have been a result of the full model's small vocabulary, which had to be artificially limited for technical reasons. Simulation 2 tested this hypothesis using a model with simpler language knowledge but a larger vocabulary, and provided some evidence that alleviating this limitation helps to make the frequency effects more monotonic. In addition, by demonstrating that a model that cannot make use of predictability information shows smaller apparent predictability effects, Simulation 2 demonstrated that the predictability effects obtained for the full model were not likely to have been merely an artifact of the correlation between word predictability and other variables such as word length.

Taken together, these results demonstrate that the rational reading framework can produce reasonable effects of word predictability and frequency on four aggregate measures of eye movement behavior: first fixation durations, gaze durations, skip rates, and refixation rates. While the quantitative fit to human data is not perfect, the fact that it is such a good match is striking given that we fit no free parameters to human data, except the conversion of timesteps to milliseconds – a parameter that all timestep-based models must include. (In future work, determining the model's best possible fit to human data will require tuning the only two other truly free parameters of our model – the agent's value of speed relative to accuracy γ and the overall visual input quality Λ.) Instead of being selected to maximize the model's fit to human data, the policy parameters α and β of our model were set to values that optimized the efficiency with which the model identified the text, given the agent's particular goal function. Future work must be done to explore the predictions of our model for a wider range of eye movement phenomena observed in reading, extending our analyses of the model's behavior both with more dependent measures, such as character landing positions within words and regressive saccades, and with more independent variables, such as word length.

Acknowledgments

This research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to K. B. and by NSF grant 0953870 and NIH grant R01-HD065829, both to R. L.

References

Allauzen, C., Riley, M., Schalkwyk, J., Skut, W., & Mohri, M. (2007). OpenFst: A general and efficient weighted finite-state transducer library. In *Proceedings of the Ninth International Conference on Implementation and Application of Automata, (CIAA 2007)* (Vol. 4783, p. 11-23). Springer.

Bicknell, K., & Levy, R. (2010a). Rational eye movements in reading combining uncertainty about previous words with contextual probability. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1142–1147). Austin, TX: Cognitive Science Society.

Bicknell, K., & Levy, R. (2010b). A rational model of eye movement control in reading. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics ACL* (pp. 1168–1178). Uppsala, Sweden: Association for Computational Linguistics.

Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science*, *21*, 366–371.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813.

Engel, G. R., Dougherty, W. G., & Brian Jones, G. (1973). Correlation and letter recognition. *Canadian Journal of Psychology*, *27*, 317–326.

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*, 487–490.

Hyönä, J., Niemi, P., & Underwood, G. (1989). Reading long words embedded in sentences: Informativeness of word halves affects eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 142–152.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Legge, G. E., Hooven, T. A., Klitz, T. S., Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: new insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219–2234.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 21086–21090. (Correction in: *Proceedings of the National Academy of Sciences of the United States of America, 107*, 5260)

Ng, A. Y., & Jordan, M. (2000). PEGASUS: A policy search method for large MDPs and POMDPs. In *Uncertainty in Artificial Intelligence, Proceedings of the Sixteenth Conference* (pp. 406–415).

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, *116*, 207–219.

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1142–1158.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

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

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*, 1270–1281.

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