Towards a Data-Driven Model of Eye Movement Control in Reading

Mattias Nilsson
Department of Linguistics and Philology
Uppsala University
mattias.nilsson@lingfil.uu.se

Joakim Nivre
Department of Linguistics and Philology
Uppsala University
joakim.nivre@lingfil.uu.se

Abstract

This paper presents a data-driven model of eye movement control in reading that builds on earlier work using machine learning methods to model saccade behavior. We extend previous work by modeling the time course of eye movements, in addition to where the eyes move. In this model, the initiation of eye movements is delayed as a function of on-line processing difficulty, and the decision of where to move the eyes is guided by past reading experience, approximated using machine learning methods. In benchmarking the model against held-out previously unseen data, we show that it can predict gaze durations and skipping probabilities with good accuracy.

1 Introduction

Eye movements during reading proceed as an alternating series of fixations and saccades with considerable variability in fixation times and saccade lengths. This variation reflects, at least to some extent, language-related processes during reading. Much psycholinguistic research, therefore, relies on measures of eye movements in reading to gain an understanding of human sentence processing. Eye tracking recordings are routinely used to study how readers' eye movements respond to experimental manipulation of linguistic stimuli (Clifton et al., 2007), and corpus-based analysis of eye-tracking data has recently emerged as a new way to evaluate theories of human sentence processing difficulty (Boston et al., 2008; Demberg and Keller, 2008).

More detailed accounts of the workings of the eye movement system during reading are offered by computational models of eye movement control (see Reichle (2006b), for an overview of recent models). These models receive text as input and produce predictions for the placement and duration of fixations, in approximation to human reading behavior. Because eye movements in reading rely on a coupled cognitive-motor system, such models provide detailed accounts for how eye movements are controlled both by on-line language processing and lower-level motor control. Current models such as E-Z Reader (Reichle, 2006a; Pollatsek et al., 2006; Reichle et al., 2009) and SWIFT (Engbert et al., 2002; Engbert et al., 2005) account for numerous of the known facts about saccade behavior in reading. This includes word frequency and predictability effects on fixation times, word skipping rates, and preview and spillover effects.

A recent approach to eye-movement modeling, less tied to psychophysiological assumptions about the mechanisms that drive eye movements, is to build models directly from eye-tracking data using machine learning techniques inspired by recent work in natural language processing. Thus, Nilsson and Nivre (2009) show how a classifier can be trained on authentic eye-tracking data and then used to predict the saccade behavior of individual readers on new texts. Methodologically this differs from the standard approach in computational modeling of eye movement control, where model parameters are often fitted to data but model predictions are not evaluated on unseen data in order to assess the generalization error of these predictions. Without questioning the validity of the standard approach, we believe that the strict separation of training data and test data assumed in machine learning may provide additional insights about the properties of these models.

The model of Nilsson and Nivre (2009) is based on a simple transition system for saccadic movements, a classifier that predicts where to fixate next and a classifier-guided search algorithm to simulate fixation sequences over sentences.

Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics, ACL 2010, pages 63–71,
Uppsala, Sweden, 15 July 2010.

One obvious limitation of the model proposed by Nilsson and Nivre (2009) is that it does not at all capture the temporal aspects of eye movement behavior. Thus, for example, it says nothing about when eye movements are initiated or when the decision of where to fixate next is made during fixations. In this paper, we try to overcome this limitation by placing the machine-learning approach in a broader psychological context and detail a model that also accounts for the timing of fixations. More precisely, we present a model of the time course of eye movements, where saccade timing is driven by on-line language processing and where-decisions are driven by the experience readers have built up through years of reading practice.[1]

It is not our intention in this paper to present a full-fledged model of eye movement control in reading. The model is limited in scope and does not address certain important aspects of eye movement control, such as within-word fixation locations, refixations and regressions triggered by higher-order processing. In addition, the linguistic features influencing timing (when-decisions) and target selection (where-decisions) are restricted to the basic variables word length and frequency. In this way, we hope to provide a baseline against which richer models of language processing can be evaluated.

The rest of this paper is structured as follows. Section 2 provides a brief background on what is known about the time course of eye movements during reading. Here we introduce some common notions that will be used later on. In section 3, we first give an overview of the model and then describe its component processes and how these processes interrelate. In section 4, we present an experimental evaluation of the model using data from the English section of the Dundee corpus (Kennedy and Pynte, 2005). Section 5 contains our conclusions and suggestions for future research.

2 The Timing of Eye Movements

The average fixation duration in reading is about 250 ms, and most fixations last between 200-300 ms, although they may range from under 100 ms to over 500 ms for a given reader (Rayner, 1998). Because eye movements are a motor response requiring preparation before execution, they are initiated well before the end of the fixation. Hence, there is a *saccade latency* of about 150-200 ms from the time when a saccade is first initiated until the eye movement is actually executed (Becker and Jürgens, 1979; McPeek et al., 2000). Once the eye movement is executed, it takes about 25-45 ms before the eyes are fixated on a new word again, depending on the length of the movement.

Given an average saccade latency of about 150-200 ms, and an average fixation duration of 250 ms, it seems clear that eye movements are often initiated within the first 100 ms of a fixation. However, as Reichle notes (Reichle et al., 2003), since the time it takes to identify words is on the order of 150 - 300 ms, this suggests that there is not enough time for language processes to have any direct on-line influence on eye movements. One key observation to explain language influences on eye movements, however, is the finding that readers often start processing upcoming words before they are fixated. Studies on *parafoveal preview* show that the amount of time spent fixating a word depends, among other things, on how much parafoveal preview of the word is available prior to the word being fixated (Balota et al., 1985; Pollatsek et al., 1992).

A further finding supporting the assumption that language processes can have an early effect on eye movements comes from the disappearing text studies (Rayner et al., 1981; Rayner et al., 2003). In these studies, words become masked or disappear at a certain point during the fixation. Despite this, a word need only be on display for 50-60 ms in order for reading to proceed quite normally. More importantly, the time the eyes remain fixated after a word disappears depends on the frequency of the word. Readers remain fixated on low-frequency words longer than on high-frequency words, even though the word that was fixated has actually disappeared. In summary, these studies suggest that there is a robust word frequency effect in reading as early as 60 ms after the onset of the fixation.

3 A Model of Eye Movement Control

3.1 General Overview

The model we develop takes the basic time constraints associated with language processing and motor control as a starting point. This means that our model is driven by estimates of the time it

[1]This view of where-decisions being driven by experience is similar in spirit to some earlier theories of saccade target selection in reading, such as the probabilistic account of word skipping proposed by Brysbaert and Vitu (1998).

takes to process words, plan an eye movement, execute a saccade etc. In line with cognitive control models of eye movements in reading, such as E-Z Reader, we assume that the cognitive processing of words is the “engine” that drives eye movements. That is, eye movements are initiated in response to on-line language processing. Unlike E-Z Reader, however, we do not presume a two-stage lexical process where the completion of a certain hypothesized first stage triggers an eye movement.[2] Instead, when the eyes move to a new word, an eye movement is initiated after some delay that is proportional to the amount of cognitive work left on the word. Furthermore, in contrast to E-Z Reader we assume that saccade initiation is decoupled from the decision of where to move the eyes. In E-Z Reader, the initiation of a saccade program is in effect a decision to start programming a saccade to the next word. Here, instead, the target for the next saccade can be any of the words in the forward perceptual span. Another related difference, with respect to previous cognitive control models, is that we assume that the decision of where to move the eyes is not directly influenced by on-line language processing. Instead, this decision is governed by an autonomous routine, having its own dynamics automated through years of reading experience. This experience is approximated using machine learning methods on authentic eye tracking data.

The model is defined in terms of four processes that we assume are operative during reading: lexical processing (*L*), saccade initiation delay (*D*), motor programming (*M*), and saccade execution (*S*). These processes are defined in terms of a set of parameters that determine their duration. Once an ongoing process ends, a subsequent process is initiated, for as long as reading continues. As is commonly assumed in most models of eye movement control, language-related processes and motor control processes can run in parallel. We will use the notation w_i to refer to the ith word in a text $w_1, \ldots, w_n$ consisting of n words, and we will use subscripted symbols L_i, D_i, M_i and S_i to refer to the lexical processing, the saccade initiation delay, the motor programming, and the saccade execution associated with w_i.

In the following four subsections, we outline these processes in detail and discuss the general assumptions underlying them. We then conclude this section by summarizing how the processes dynamically interact to produce eye movement behavior.

3.2 Lexical Processing

The time needed to process individual words in reading is certain to depend on numerous factors related to a person's prior reading experience, word-level properties such as length and frequency, and higher-order language processes such as syntactic and semantic processing. However, since our goal in this paper is to validate a simple model, with as few parameters as possible, we make the simplifying assumption that the processing time of a word can be approximated by its length (number of characters) and its frequency of occurrence in printed text. In particular, we assume that the mean time required for processing a word w_i is a linear function of its length and the natural logarithm of its frequency:[3]

$$t(L_i) = b_0 + b_1 \,\text{length}(w_i) - b_2 \,\ln(\text{freq}(w_i)) \quad (1)$$

In equation 1, b_0 is the intercept representing the base time needed to process a word while b_1 and b_2 are the respective slopes for the effect of length and frequency on the base processing time. Again, we stress that equation 1 is by all accounts an oversimplification. Thus, for example, it does not take into account any higher-level top-down influence on processing time.

Still, we believe equation 1 provides a reasonable first approximation. A large part of the variance in measures of reading time can be accounted for by word frequency and word length. At any rate, our simple assumption with respect to processing time represents a methodological decision rather than a theoretical one. We want to keep the model as simple as possible at this stage, and later explore the effect of including variables related to higher-order processing.

Once the time interval $t(L_i)$ has passed for a given word w_i, lexical processing begins on the next word. Thus, the completion of $t(L_i)$ results in the initiation of L_{i+1}. Because the processing of the next word does not start until the processing of the current word is finished, lexical processing

[2] In E-Z Reader, the first stage of lexical processing is an early estimate of the word's familiarity that provides the signal to the eye movement system that lexical access is imminent and that a saccade should be planned.

[3] We use the logarithm of word frequency because human response times, in lexical decision tasks for instance, are linearly related to the natural logarithm of word frequency (Balota and Chumbley, 1984).

proceeds serially and no more than one word is processed at any given time.

3.3 Saccade Initiation Delay

When the eyes move to a new word w_i, a motor program is initiated after some time. We assume that the time when a motor program is initiated depends on the processing difficulty of the fixated word w_i. In particular, the signal to initiate a saccade is deferred in proportion to how much processing remains on w_i, or put differently, in proportion to how much work remains to be done on that word. This general routine serves to prevent the control system from making over-hasty saccades to new words. The length of the saccade initiation delay $t(D)$ is proportional to the remaining processing time of word w_i at fixation onset:

$$t(D_i) = d\,(t(L_i) - t(E_i)) \qquad (2)$$

where d is a free parameter representing a proportion, $t(L_i)$ is the lexical processing time for the fixated word, and $t(E_i)$ denotes the interval of time that has elapsed since the initiation of $t(L_i)$. More difficult words are associated with longer processing times and thus cause later initiation of saccade programs and therefore also longer fixation durations. The free parameter d defines a proportion taking values in the range $[0, 1]$. The extremes of this range can be interpreted as follows. If d is set equal to 0, a new saccade program is initiated immediately upon a new fixation. If d instead is set equal to 1, the saccade program starts only after the fixated word has been fully processed. More generally, a change of the value of this parameter can be understood as a change of the amount of cognitive influence on fixation durations. The higher its value, the more cognitive work must be carried out before a new saccade program is started. Once the time interval $t(D)$ has passed, the planning of a new eye movement starts, i.e., a motor program, M, is initiated.

3.4 Motor Programming

The time needed to plan and initiate an eye movement defines the saccade latency, or motor programming time $t(M)$. We assume that the duration of this period is given by the free parameter m:

$$t(M_i) = m \qquad (3)$$

The following is worth noting. Some influential research suggests that motor programming is completed in two stages (Becker and Jürgens, 1979). The first of these being a labile stage during which a planned saccade can be canceled, e.g., in favor of another saccade target. The second stage, closer in time to the execution of the saccade, is non-labile and once entered, a saccade underway can no longer be modified or canceled. This division between labile and non-labile stages of motor programming is sometimes implemented in computational models, for example in E-Z Reader and SWIFT. For now, however, our model does not operationalize the notion of saccade canceling and thus makes no useful distinction between labile and non-labile stages of motor programming. Our only assumption with respect to these different stages of motor programming is that their respective durations sum up to m.

An important function of motor programming in our model, however, is to select a target for the saccade. Before discussing how this is achieved we should point out that we make no claim as to how much time of motor programming is consumed by target selection. It is only presupposed that saccade target selection, in the normal course of events, is initiated as soon as there is a decision to make an eye movement (i.e., when motor programming starts), and that, whatever time remains of motor programming once a target is selected, this time is spent on preparation of the physical movement to the selected target. Once motor programming is finished, a saccade S is executed to the target.

Following Nilsson and Nivre (2009), we treat target selection as a classification task. In practical terms, this means that we train a classifier to predict the most likely eye movement following any fixation. An instance to be classified consists of a feature vector encoding feature information over the current fixated word and words in the immediate context. Given such feature representations and training data obtained from eye-tracking recordings, essentially any standard machine learning algorithm can be applied to the classification task. The type of learning algorithm that performs best on this task is, however, unknown. Rather than speculate, we suggest that this is a question for further research.

The remaining assumptions we make are as follows. First, because there is a sharp drop-off in acuity of the human eye around the point of fixation, the number of words that can be discriminated in parafoveal vision on a given fixation is limited to a few. Therefore, it is reasonable to assume that the potential targets for a saccade on any given fixation are limited to the words available within the range of effective vision. [4] This is supported empirically by the fact that the great majority of outgoing saccades tend to land in one of the three words that follow the current fixation. Moreover, we assume that for these potential targets, only rather coarse, visual information, such as a gross appreciation of their length, can be extracted on any given fixation. The reason for this is that target selection generally occurs relatively early on in a fixation, at a time when only low-level visual information can reasonably be gleaned from the parafovea.

Secondly, we reason that target selection reflects an autonomous process that has been automated, through years of practice, to *progress* through the text and select targets in the default reading direction. Hence, the possible targets for target selection, as construed here, is limited to the targets within the forward field of effective vision. As a consequence, words to the left of the current fixation are not fixated as a result of target selection.

Finally, we assume that target selection by default is a mechanical routine, insensitive to ongoing lexical processing. In the general case, then, the decision of where to move eyes is made independently of processing considerations. Motor programs in general, however, may sometimes override the default target selection mechanism and be initiated, not in order to *select* a new target, but to *correct* for situations where motor control and ongoing language processing are threatening to desynchronize. Such a corrective program may be initiated, for instance, if a saccade is executed to word$_i$ but lexical processing has not yet completed on word$_{i-1}$, and so more lexical processing of word$_{i-1}$ is needed before moving on. In this case, a corrective motor program is initiated to word$_{i-1}$, subsequently resulting in a regression to that word. In this way, corrective motor programs serve to synchronize the eyes with the current processing stream and for that reason they always target the word being processed. Moreover, because corrective saccade programs are launched with a fixed target, they do not trigger target selection during motor programming.

[4]The effective visual field (the perceptual span) extends about four characters to the left and 15 characters to the right of the fixation for normal readers of left-to-right orthographies (Rayner, 1998).

3.5 Saccade Execution

The time to execute a saccade $t(S)$ is determined by the free parameter s:

$$t(S_i) = s \tag{4}$$

Once a saccade has been executed, the position of the eyes shifts to a new word and thus, in the normal course of events, a new motor program is initiated after $t(D_i)$. However, sometimes a saccade is made ahead of the current processing stream, because, as noted earlier, a word needs not be fully processed before a saccade is executed to another word. Likewise, a saccade may sometimes be executed to a word that has already been fully processed, because target selection is an autonomous process, not influenced by ongoing processing. In these situations, corrective saccade programs are initiated. Since corrective saccade programs serve only to rapidly coordinate the eyes and the current processing stream, we assume that they can be initiated immediately and hence that they are not subject to saccade initiation delay.

3.6 Eye Movement Control

Having defined the respective component processes, we now consider how these processes are coordinated to model eye movement control. Lexical processing is always running in parallel with the processes controlling saccade initiation delay, motor programming and saccade execution, which are executed in sequence. A simulation of reading is started by initiating lexical processing of the first word (L_1), and the saccade initiation delay for the first word (D_1) (i.e., the first word is fixated). Whenever one of the running processes terminates, new processes are initiated in the following way:

- If L_i terminates, initiate L_{i+1}.
- If D_i terminates, initiate M_i and select new fixation target w_j.
- If M_i terminates, initiate S_i.
- If S_i terminates and the ongoing lexical process is L_j:
  - If $i = j$, initiate D_i.
  - If $i \neq j$, initiate M_j and set fixation target to w_j

The simulation terminates when all words have been lexically processed.

4 Experimental Evaluation

4.1 Experimental Setup

In order to estimate the performance of the model described in the previous section, some experiments were performed using data from the English section of the Dundee corpus (Kennedy and Pynte, 2005).

In most evaluations of eye movement control models, the model parameters are fitted against one and the same corpus by searching the parameter space to find the set of parameter values that best simulates the observed data. This approach makes it somewhat hard to appreciate how well a given model generalizes to new, previously unseen data. A more stringent evaluation, which affords an assessment of the generalization error of model predictions, is to set the model parameters on some portion of the data and then test the model on another held-out portion. The results we report in this paper were obtained this way.

The Dundee corpus that was used in these experiments contains the eye tracking records of ten subjects reading editorials from The Independent, a UK broadsheet newspaper. The data consist of 20 texts that were read by all subjects, and close to 2400 sentences. We divided these texts into three sets: the first 16 for training (1911 sentences), 17-18 for model development and validation (237 sentences), and the last two texts, 19-20, for blind testing of the model (231 sentences). Model parameters were fitted using only the training and validation set, prior to evaluating the model on the held-out test set.

Next we discuss how training was performed, both in terms of the training of the classifier for target selection and in terms of the estimation of the model's process parameters on the training data. Before presenting the results, we also discuss some standard practice in benchmarking models of eye movement control.

4.2 Training the Classifier

We used the transition-based model outlined by Nilsson and Nivre (2009) in combination with logistic regression for training the target selection classifier. The classifier was trained on a restricted number of features defined over words in the fixation context. The feature model we used for these experiments included information about the word length of the current fixation and upcoming words, as well as some historical information about recently made eye movements. The history of previous eye movements was represented in terms of the saccade distance (measured in number of words) that led up to recently made fixations (including the current fixation). In this way, the feature model contained information about, for instance, whether the saccade that led up to the current fixation skipped a word or two.

In contrast to Nilsson and Nivre (2009) we did not train one model for each individual subject in the corpus. Instead, we trained a single multiple-subject classifier on all ten readers in the training set. The performance of this classifier was assessed in terms of how well, on average, it predicted the observed saccade targets for any given reader on the development set. Moreover, in line with the assumption that target selection is restricted to a limited number of candidate words in the forward visual field, the classifier was trained to select one of the three words following any fixation as the target for a saccade. This cross-subject classifier achieved an average prediction accuracy of 72% on the development set.

4.3 Estimating Model Parameters

Because the model's process parameters can not be directly estimated from eye tracking data they need to be approximated in other ways. The values for the intercept and slope parameters for lexical processing time $t(L_i)$ were obtained by fitting a linear regression of gaze duration on logarithmic word frequency and word length on the training data. The assumption that the gaze duration on a given word reflects the time required to process the word is necessarily an oversimplification but is sometimes used in eye movement modeling. A number of studies indicate that it is indeed a reasonable approximation (Engbert et al., 2002; Pollatsek et al., 2006).

The value for the parameter d in the equation for $t(D_i)$ was selected based on a simple parameter search over the training data. The best fitting value was assessed by calculating the root mean square error between predicted and observed values for gaze durations for different values of d ranging from 0 to 1 in 0.1 increments, while keeping other parameter values unchanged. To keep things simple, the parameters that determine the mean duration of motor programming, m, and saccade execution, s, were fixed at 200 ms, and 25 ms, respectively. These values are in good agreement with

Parameter	Interpretation	Value
b_0	Intercept: base lexical processing time (ms)	165.5
b_1	Slope: effect of length on lexical processing time (ms)	13.5
b_2	Slope: effect of frequency on lexical processing time (ms)	3.2
d	Proportion of lexical processing time (determines saccade initiation delay)	0.5
m	Mean motor programming time (ms)	200
s	Mean saccade execution time (ms)	25

Table 1: Model parameters, their interpretations and values, as estimated during training.

estimated values in experimental studies. Table 1 lists the model's six process parameters and their values, obtained prior to testing the model.

4.4 Benchmark Evaluation

Models of eye movement control in reading are typically benchmarked against a set of word-based dependent eye movement measures which are averaged across subjects. Two such measures are gaze duration and probability of skipping. Gaze duration is defined as the sum duration of all fixations on a word prior to any saccade leaving the word during first-pass reading. Probability of skipping is simply the mean probability (across subjects) that a given word is skipped (not fixated) during first-pass reading.

Because word frequency effects on eye movements during reading are robust and well-documented, one common benchmark practice is to evaluate models with respect to their capability of reproducing word frequency effects on fixation times and fixation probabilities. Typically, averages of word-based measures are then broken down into word-frequency classes. This is a fairly simple way to see how well a given model can predict observed means for measures such as gaze duration and skipping probability for words of different frequency classes. The results we report are presented this way. We used frequency estimates based on word occurrences in the written part of the British National Corpus (BNC). Frequencies were normalized to occurrences per million words and then divided into five frequency classes, as suggested by Reichle et al. (1998).

In addition to the model we have outlined so far, we also present results for two alternative versions. These models differ from the one we have discussed only in positing a simpler function for lexical processing time. The alternative versions model lexical processing time only as a linear function of either word length or logarithmic word frequency. Hence, we fitted two separate simple linear regressions of gaze duration first on word length, and then on logarithmic word frequency. The regression coefficient and slope were estimated to 132.5 and 16 for the model based on word length, and 284 and -11 for the model based on frequency.

4.5 Results and Discussion

Table 2 shows the observed (empirical) and predicted (simulated) values of gaze durations and skipping probabilities for each of the five word frequency classes, both on the development set and on the held-out test set. M_1 and M_2 represent the versions of the model in which lexical processing time is a linear function of word length, and word frequency, respectively. M_3 represents the version of the model where lexical processing time is a linear function of both variables.

The results show that all three models, on the development set as well as on the test set, are able to reproduce the most important aspect of the observed data, namely, that mean gaze durations decrease and mean skipping probabilities increase with increasing word frequency. Overall, M_3 performs better than the two other models in predicting this relationship. The model based only on word length, M_1, performs worse than the other two models. This is mainly due to the poor performance of this model in simulating the proportions of skipped words in the upper frequency classes 4 and 5. In comparison to both M_2 and M_3, M_1 seriously underestimates the observed skipping probability for words belonging to these frequency classes, on both development and test data.

With respect to gaze duration alone, the three models perform similarly, although M_3 provides a somewhat better fit on both data sets. The models generally predict longer gaze durations than the observed means, except for the most low-frequent words. In particular, gaze durations for higher-frequency words (class 4 and 5) are prolonged compared to the means, giving an overall nar-

	Gaze duration								Probability of skipping							
	Development				Test				Development				Test			
Frequency class	Observed	M_1	M_2	M_3	Observed	M_1	M_2	M_3	Observed	M_1	M_2	M_3	Observed	M_1	M_2	M_3
1	290	282	280	285	286	278	280	284	0.17	0.15	0.18	0.13	0.16	0.14	0.19	0.14
2	257	271	259	272	261	273	260	275	0.19	0.18	0.20	0.16	0.19	0.15	0.22	0.17
3	229	254	252	249	235	257	254	252	0.24	0.19	0.24	0.20	0.22	0.19	0.25	0.20
4	208	240	238	237	210	244	238	237	0.52	0.23	0.36	0.43	0.53	0.24	0.34	0.40
5	198	238	236	228	195	239	237	230	0.65	0.34	0.51	0.54	0.67	0.32	0.52	0.51

Table 2: Observed and predicted values of Gaze Durations (ms) and Skipping Probabilities on development and test set for five frequency classes of words. M_1: $t(L_i) = b_0 + b_1\text{length}(w_i)$, Root mean square error on development set = 0.48, Root mean square error on test set = 0.52; M_2: $t(L_i) = b_0 - b_1 \ln(\text{freq}(w_i))$, Root mean square error on development set = 0.33, Root mean square error on test set = 0.35; M_3: $t(L_i) = b_0 + b_1\text{length}(w_i) - b_2 \ln(\text{freq}(w_i))$, Root mean square error on development set = 0.21, Root mean square error on test set = 0.26; Frequency range: 1:1-10, 2:11-100, 3:101-1000, 4:1001-10000, 5: 10001+

rower range of mean values for the five frequency classes.

The overall performance of each model, M_1, M_2 and M_3 was estimated by calculating the root mean square error (RMSE) between the mean observed and predicted gaze durations and probabilities of skipping. The errors were normalized as described in Reichle et al. (1998). In comparing the results for both development and test data, the best overall fit is provided by M_3 on the development set, giving an RMSE of 0.21 (smaller values indicate better fit). The fit for the same model drops to 0.26 when evaluated on the held-out test data.

To provide some basis for comparison, the earliest version of E-Z Reader (Reichle et al., 1998) which was fitted to the same dependent measures, had an RMSE of 0.145. It is important to point out, however, that this result was based on fitting the model parameters to a single sentence corpus of 48 sentences designed for experimental purposes. This corpus contained relatively short (8-14 words) isolated sentences without any connecting discourse. More generally, as noted by Reichle et al. (2009), RMSD values lower than 0.5 provide fits that are reasonably close to the observed means. By this standard, the model M_3 performs rather well in simulating the observed data. Moreover, this version of the model provides the most realistic estimates of the time it takes to identify words. Thus, for example, the mean time to identify the most frequent word in English, "the" (frequency class 5), is estimated to be 171 ms, whereas the mean time to identify the word "repopulate", which is a low-frequency (frequency class 1) ten-letter word is estimated to be 301 ms. These estimates are in good agreement with experimental estimates, which show that word identification latencies range between 150 and 300 ms (Rayner and Pollatsek, 1989).

5 Conclusion

In this paper we built on previous work using machine learning methods to model saccade behavior in reading and we extended this work by presenting a data-driven model of eye movement control that provides detailed predictions for both *when* and *where* the eyes move during reading. The most important principles of this model are *(i)* the initiation of eye movements is delayed as a function of on-line processing difficulty, and *(ii)* the decision of where to move the eyes is driven by an autonomous routine that has become automated through years of practice in reading. The model was trained on eye movements made over a large corpus of natural text. In benchmarking the model against held-out data we showed that it is able to reproduce frequency effects on both gaze duration and skipping probability with good accuracy (RMSE = 0.26).

Looking ahead, we plan to extend the model to account for more empirical data on eye movement behavior in reading. One important step to meet this goal is to develop a more informed model of language processing. Current models of eye movement control in reading generally assume that influences from syntactic and higher-order processing occur too late in the processing stream to directly influence eye movements. This is, however, seemingly at odds with recent findings in sentence processing research showing an influence of syntactic processing difficulty on both early and late measures of eye movements in reading (Demberg and Keller, 2008; Boston et al., 2008). Hence, it is possible that a more accurate model of eye movements in reading will need to allow for syntactic processing to influence the early decisions that control the timing of eye movements. This and other issues will be addressed in future work.

References

David. A Balota and James. I Chumbley. 1984. Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and Performace*, 10:340–357.

David. A. Balota, Alexander Pollatsek, and Keith Rayner. 1985. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17:364–390.

W Becker and R Jürgens. 1979. An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19:967–983.

Marisa F. Boston, John Hale, Reinhold Kliegl, Umesh Patil, and Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Reasearch*, 2:1–12.

Marc Brysbaert and Françoise Vitu. 1998. Word skipping: implications for theories of eye movement control in reading. In Geoffrey Underwood, editor, *Eye guidance in Reading and Scene Perception*, pages 124–147. Elsevier science Ltd.

Charles Clifton, Adrian Staub, and Keith Rayner. 2007. Eye movements in reading words and sentences. In Roger van Gompel, editor, *Eye movements: A window on mind and brain*, pages 341–372. Amsterdam: Elsevier.

Vera Demberg and Frank Keller. 2008. Data from eyetracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109:193–210.

Ralf Engbert, André Longtin, and Reinhold Kliegl. 2002. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42:621–636.

Ralf Engbert, Antje Nuthmann, Eike Richter, and Reinhold Kliegl. 2005. SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112:777–813.

Alan Kennedy and Joël Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision research*, 45:153–168.

R. M. McPeek, A. A. Skavenski, and K Nakayama. 2000. Concurrent processing of saccades in visual search. *Vision Research*, 40:2499–2516.

Mattias Nilsson and Joakim Nivre. 2009. Learning where to look: Modeling eye movements in reading. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 93–101.

Alexander Pollatsek, Mary Lesch, Robin K. Morris, and Keith Rayner. 1992. Phonological codes are used in integrating information across saccades in word identification and reading. *Experimental Psychology: Human Perception and Performance*, 18:148–162.

Alexander Pollatsek, Erik Reichle, and Keith Rayner. 2006. Tests of the E-Z Reader model: Exploring the interface between cognition and eye movements. *Cognitive Psychology*, 52:1–56.

Keith Rayner and Alexander Pollatsek. 1989. *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Keith Rayner, Albert W. Inhoff, Robert E. Morrison, Maria L. Slowiaczek, and James H. Bertera. 1981. Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 7:167–179.

Keith Rayner, Simon P. Liversedge, Sarah J. White, and Dorine Vergilino-Perez. 2003. Reading disappearing text: cognitive control of eye movements. *Psychological science*, 14:385–388.

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

Erik Reichle, Alexander Pollatsek, Donald Fisher, and Keith Rayner. 1998. Toward a model of eye movement control in reading. *Psychological Review*, 105:125–157.

Erik Reichle, Keith Rayner, and Alexander Pollatsek. 2003. The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26:445–476.

Erik Reichle, Tessa Warren, and Kerry McConnell. 2009. Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16:1–21.

Eric Reichle, editor. 2006a. *Cognitive Systems Research*. 7:1–96. Special issue on models of eye-movement control in reading.

Eric Reichle. 2006b. Computational models of eye movement control in reading: Theories of the "eye-mind" link. *Cognitive Systems Research*, 7:2–3.