Framework of a probabilistic gaze mapping model for reading

Michael Carl

Abstract

One of the aims of the Eye-to-IT project is to investigate the possibility of using eye-tracking devices for detecting situations of targeted help for human translators. A prerequisite for automated assistance in human translation is the understanding and modelling of reading behaviour, the ability to follow human eye movements and to map gaze sample points — the output of eye-tracking devices — onto words and symbols fixated.

Within the Eye-to-IT project we currently use a so-called "Gaze-to-Word Mapping" (GWM) device (Špakov 2008) that first computes possible fixations from sequences of gaze sample coordinates and then maps the fixations on the words which are likely to be fixated.

This paper suggests an alternative framework of a probabilistic gaze mapping model for reading, in which fixations on textual objects are directly computed from the gaze sample points. The framework integrates various knowledge sources with the aim to compute the most likely fixations on words and symbols on the basis of the available data.

1. Introduction

The Eye-to-IT project aims at providing automatic assistance when a human translator seems to need help. The basic assumption is that translators' need for assistance can potentially be detected from their gaze path when reading the text to be translated. Whenever a translator experiences text comprehension problems, the eyes would move in an 'unusual' manner, and the system would provide help by showing word translations into the target language of the difficult passages. The Eye-to-IT project is based on a previous 'iDict' application (Hyrskykari 2006), in which readers of electronic foreign language documents are provided automatic assistance when they appear to need help. Hyrskykari (2006) points out:

> The most fundamental problems [. . .] when trying to detect deviations from the normal flow of reading can be articulated with the two main questions **where** and **when**. The third essential question is how the application should react when the probable cause or digressive reading is identified. (Hyrskykari 2006:12, original emphasis)

The same three problems also apply to the Eye-to-IT project, if translators are to be provided with automatic assistance. Much research has been carried out to understand the processes of reading, and eye movement behavior is a thoroughly studied field with respect to describing what constitutes a 'normal flow of reading'. Within the Eye-to-IT project and the iDict application three basic entities in gaze-to-word mapping are used:

1. **Gaze samples** are essentially pixel locations in terms of X/Y coordinates of the screen as produced by the eye-tracking devices.

2. **Fixations** are sets of gaze samples, expressed in terms of X/Y coordinates which correspond to the center of a set of gaze samples. Fixations are computed as an average of the sample sets.

3. **Textual objects** are possible screen locations on which fixations may take place. In the GWM tool these are sequences of characters separated by blanks (i.e. words), but in principle they may also be smaller or larger units.

According to Salvucci and Goldberg (2000:71), "fixation identification is an inherently statistical description of observed eye movement behaviors" which can be detected on the basis of the velocity or the dispersion of the gaze sample points: the former due to the observation that saccades have high velocity, and the latter because movements inside fixations occur close to each other.

In applications like iDict and GWM the program must be able to map the reader's fixations to the words being read. Without providing a complexity analysis of her algorithm, Hyrskykari (2006) argues that it is computationally too expensive in a real-time application to map every gaze position directly on the text objects. She therefore first identifies fixations on the basis of the dispersion of raw gaze data samples. In a second step, the fixations, an already reduced and filtered amount of information, are mapped on the textual objects. In this algorithm, the second step is the difficult task, where fixation-to-word mapping has to compensate for vertical and horizontal inaccuracies as well as calibration drift. In order to compute such mappings Hyrskykari (2006) introduces a number of heuristics:

- lines are read from left to right
- almost every content word is fixated upon at least once
- about 10 to 15 percent of saccades are regressions, often to the same line but sometimes to the lines previously read
- at the end of a line, the reading point is transferred to the beginning of the next line

Within the Eye-to-IT project, we are facing several shortcomings of this procedure, partially due to the fact that reading with different purposes also produces different reading behavior. Jakobsen and Jensen (2007, and in this volume) show that far more regressions occur when reading with the purpose of translation; here the eye-gaze patterns are very different from usual reading patterns. Hence, there is a need be able to easily adapt eye-gaze mapping models to particular situations, readers or contexts without having to invent and adapt new heuristics.

We therefore propose a general probabilistic gaze-mapping model which can be trained on available hand-corrected gaze-mapping data so as to produce special purpose gaze-to-word mapping models. These specialised models estimate fixation positions and durations based on the knowledge of the reading task — as provided by a training corpus — and the knowledge of the textual objects, thereby collapsing fixation detection and fixation-object mapping into one computational process consisting of several probabilistic factors.

In section 2. the parameter setting of the probabilistic model is outlined. Examples are given of parameter distributions computed from a text of 100 words that was read by four different translators and which represents an overall reading time of approximately four minutes.

In section 3., the model is evaluated through a re-estimation of the GWM gaze-mapping and compared to a gold standard. Finally further perspectives are pointed out for the future development of the model.

2. A probabilistic model of fixation for reading

We suggest computing fixations from gaze sample points in which account is taken of the distribution of textual objects on the screen. Given a set of gaze sample points $S_{1\ldots m}$ and the fragmentation of the screen into textual objects $O_{1\ldots n}$ we aim at estimating the probability of a fixation F_j on object O_j according to equation (1):

$$P(F_j \mid S_{1\ldots m}) = \underbrace{\sqrt[m]{\Pi_{i=1}^{m} P(O_j \mid S_i)}}_{\text{SO-distance}} * \underbrace{P(O_j \mid O_{j-1})}_{\text{O}^2\text{-transition}} * \underbrace{P(F_j \mid O_j)}_{\text{OF-length}} \tag{1}$$

The underlying assumption is that textual objects represent the only possible fixation locations in a reading task, and that the probability of a fixation can be broken down into independent factors, each of which contributes to the object-fixation probability. The model in equation (1) assumes three factors:

1. The Sample/Object (SO) distance indicates the probability that sample(s) S_i is part of a fixation on the textual object O_j.

2. The Object/Object (O^2) distance indicates the probability of a saccade from object O_{j-1} to object O_j.

3. The Object-Fixation (OF) length indicates the probability of the expected fixation length on object O_j.

The following sections illustrate how these probabilities may be computed based on a collection of manually-adjusted gaze-to-word mappings.

2.1 SO-distance: how far away may a sample be located from a fixated object

The sample-object distance SO provides the probability that a given sample S_i 'belongs' to object O_j. This probability takes into consideration characteristics of the textual objects and the distance $dist(O_j, S_i)$ between the object O_j and the pixel position of S_i, as in equation (2).

$$P(O_j \mid S_i) = P(O_j \mid dist(O_j, S_i)) = \frac{Cnt(size(O_j), dist(O_j, S_i))}{Cnt(dist(O_j, S_i))} \tag{2}$$

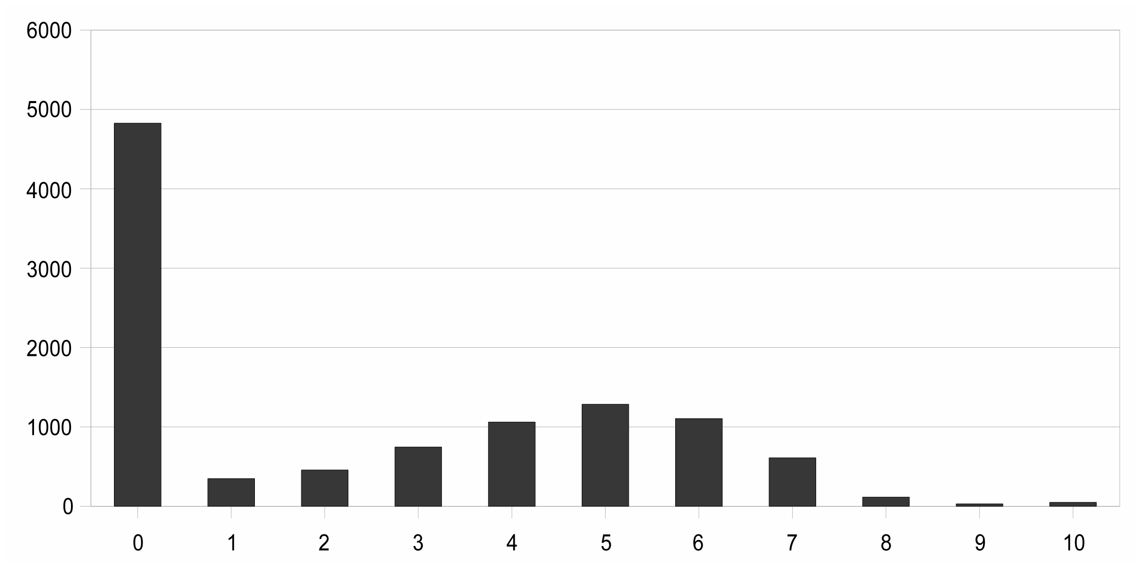

Figure 1. Distribution of sample-object distances $dist(O_j, S_i)$ in the data according to equation (3). Note that distances on the X-axis are on a logarithmic scale. The classes 1 to 7 represent sample-object distances of up to 1, 2, 4, 7, 12, 19 and 32 pixels respectively. Most samples are in the textual objects' box (class 0) while there is a peak around $\sqrt{e^4}$ to $\sqrt{e^5} = 7$ to 12 pixels distance between the sample points and the object's outer border.

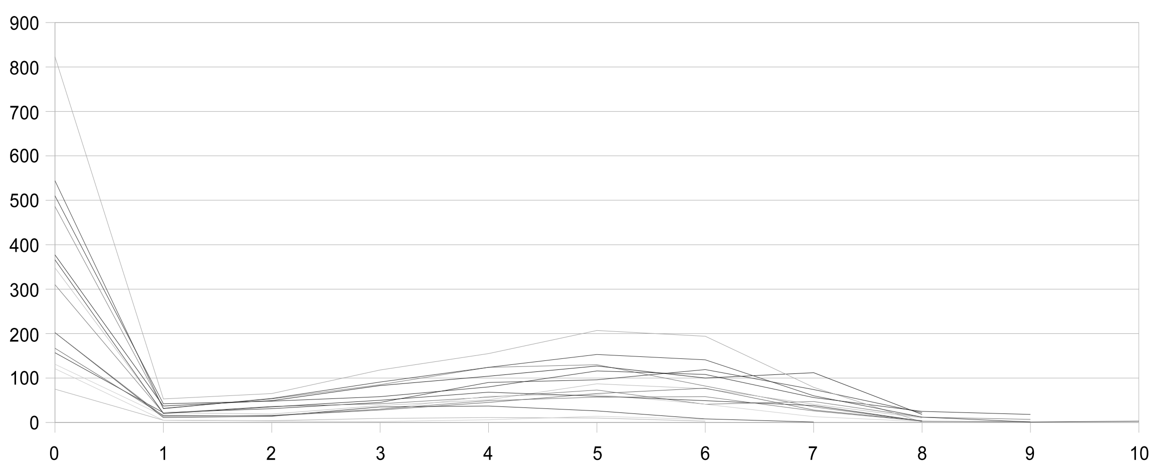

Figure 2. Distribution of object-size/sample-object distance $\{size(O_j), dist(O_j, S_i)\}$. Each line represents a different object-size $size(O_j)$. The distribution of sample-object distance for each object size resembles, and is similar to, the graph shown in figure 1.

In the current model we set the marginal probability $P(O_j)$ equal to the probability of the size of the textual object, taking as characteristic feature of O_j only its size in terms of characters. The size $size(O_j)$ therefore equals the number of characters of O_j.

The distance between the object and the sample is a number ≥ 0 indicating how far away the sample point is from the outer border of the object. The distance $dist(O_j, S_i) = 0$ if the sample point is inside the object box; if not, it is the natural log of the square distance, computed as in equation (3):

$$\begin{array}{lll} \text{if} & (S_i \text{ inside } O_j) & : dist(O_j, S_i) = 0; \\ \text{else} & & : dist(O_j, S_i) = log_e((Y_{O_j} - Y_{S_i})^2 + (X_{O_j} - X_{S_i})^2 + 1); \end{array} \quad (3)$$

where X_{S_i} and Y_{S_i} represent the X and Y coordinates of the gaze sample points S_i; and X_{O_j} and Y_{O_j} are the X and Y coordinates on the outer border of object O_j, which are closest to the sample point S_i. Figures 1 and 2 plot these distributions for the sample set. The probabilities are approximated by counting ($Cnt(\cdot)$) the number of occurrences in the data.

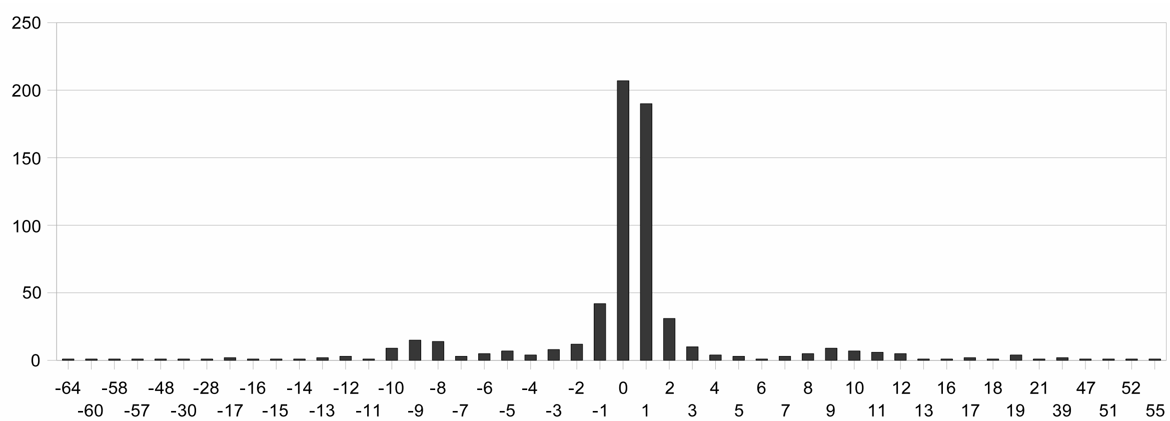

Figure 3. Distribution of saccade lengths between successively fixated objects. The X-axis represents saccade length in terms of characters/10. Most saccades are progressive and occur between objects of up to 10 or 20 characters distance. Small peaks are at a distance of $\pm$ 80 to 100 characters. Many of the contributing fixations in these peaks are likely to be due to fixation mapping errors in the data, where fixations were erroneously mapped on a textual object in the line above or below the currently read line.

2.2 O^2 distance: how large are distances between successively fixated object

This factor indicates the probability of a saccade occurring between object O_{j-1} and object O_j:

$$P(O_j \mid O_{j-1}) = P(O_j \mid dist(O_i, O_{j-1})) = \frac{Cnt(size(O_j), dist(O_j, O_{j-1}))}{Cnt(dist(O_j, O_{j-1}))} \tag{4}$$

Similar to SO-distance, we set the marginal probability $P(O_j)$ equal to the probability of the size of the textual object. The distance $dist(O_j, O_{j-1})$ between two textual objects O_j and O_{j-1} equals the number of characters between the first characters of the two objects divided by 10. Figure 3 shows distributions of saccade lengths between successively fixated textual objects, as retrieved from our data set.

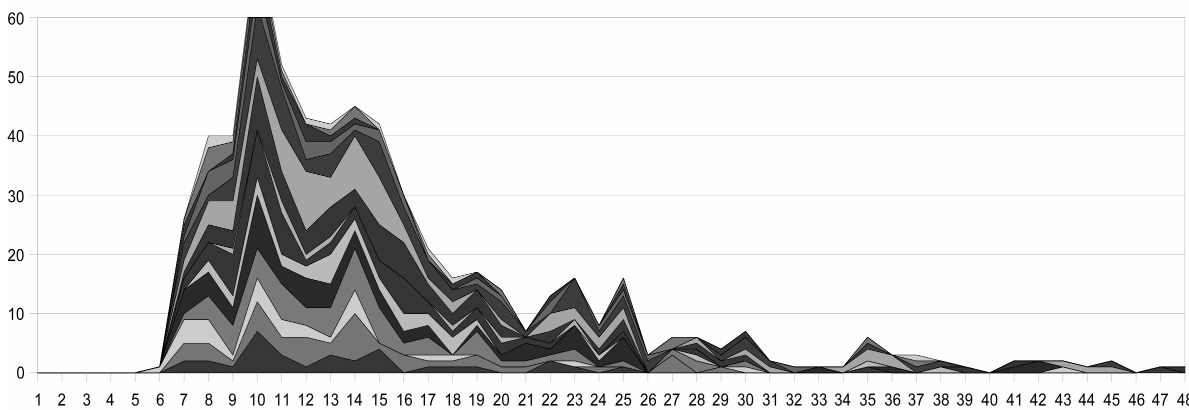

Figure 4. Distribution of object-size/fixation length $\{length(F_j), size(O_j)\}$. The X-axis shows the number of samples per fixation; the Y-axis the number of fixations counted in the data set; each color in the graph represents a different object size. Most fixations contain between 9 and 16 samples; with a sampling rate of 50 Hz, this equals fixation length of 180 to 320 ms.

2.3 OF-length: how likely is the fixation length on an object

A third factor taken into account is the probability of the fixation length on a given object O_j. Similar to equation (4), we also compute the size of O_j in terms of characters, while the length $length(F)$ is determined by the number of samples counted in fixation F. Figure 4 plots distribution of fixation length, with each sample representing 20ms.

$$P(F_j \mid O_j) = P(length(F_j) \mid size(O_j)) = \frac{Cnt(length(F_j), size(O_j))}{Cnt(size(O_j))} \tag{5}$$

3. Evaluation and outlook

This paper has presented preliminary investigations for a probabilistic framework of fixation modelling during reading. Section 2. introduced the probability model and showed figures of the distribution as obtained from a small data collection of four reading experiments using a short English text of 100 words.

Figure 5 shows the results of a comparison of the re-estimated fixations using equation (1) with a manually adjusted gold standard on the same text of 100 words. The gold standard and the statistically re-estimated fixations differ in various aspects: while manual adjustment exhibited 171 fixations, only 148 fixations were detected during statistical fixation re-estimation. Also many of the fixations have different lengths and/or different starting points in the manually-adjusted compared with the statistically re-estimated data. More than 80 fixations, i.e. 56% of the 148 statistically detected fixations are identical, showing (almost) identical fixation onset times.

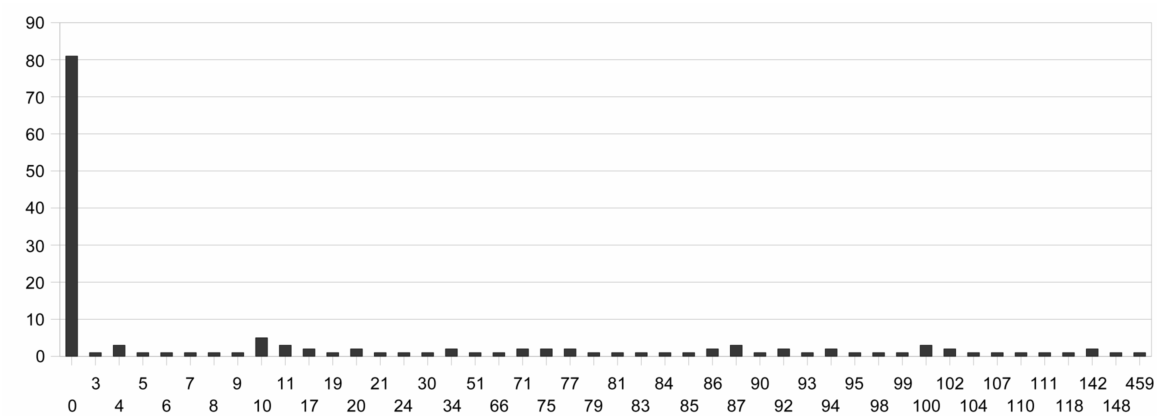

Figure 5. The graph plots the distances between the manually adjusted fixation mappings and the statistical fixation re-estimations according to equation (1). Around 56% of the gold standard and statistical fixation mappings are identical. The 44% remaining fixations are almost equally distributed at a distance ranging from 3 to 150 characters.

The relatively small number of fixations during statistical re-estimation may be due to the fact that fixations are differently computed. While dispersion-based fixation detection in GWM (Špakov 2008) only considers samples within a distance of approximately 30 pixels as belonging to the same fixation, the statistical device clusters all samples that fall inside (or close to) a textual object onto the same fixation, according to the probability model. Several successive GWM fixations which are mapped onto the same textual object would thus collapse into one fixation during statistical re-estimation.

In the future we aim at developing various aspects of the probabilistic model. In addition, we would like to evaluate it with larger data collections. In particular, the following points will be investigated:

- More factors might be required to model fixation mapping. At present, the model does not take into account the sample velocity, which might be computed as the distance between successive sample points, and which might give indications on dynamics of fixations as well as their starting points and end points.

- Granularity of the textual objects: As outlined above, the size of the textual objects determines the dispersion and possible distances of fixations. By manipulating the object size we are likely to obtain different fixation resolution and distribution. An optimum object size might be computed based on the sample point distribution that can be observed in the data.

- Properties of the textual objects: in the present analysis we have only considered the size of the textual object as its characteristic feature. However, there is much evidence that many other features play a role in eye-movement control during reading, e.g.:

  – linguistic factors: part of speech of the words, position within phrases or clauses, semantic properties of the words, etc.

  – length, familiarity, frequency, collocational information, word ambiguity, etc.

We aim at investigating whether, and if so how, these features might be taken into account in the probabilistic framework, and what their impact on accurracy and predictability for gaze-to-word mapping is.

References

Hyrskykari, A. 2006. *Eyes in Attentive Interfaces: Experiences from Creating iDict, a GazeAware Reading Aid*. Ph.D thesis, University of Tampere, Tampere.

Jakobsen, A. L. and Jensen, K. T. H. 2007. Coordination of comprehension and text production in witten and oral translation tasks. *Paper presented at AMLaP 2007*.

Salvucci, D. D. and Goldberg, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye tracking Research & Applications (ETRA'00)*, 71–78.

Špakov, O. 2008. Gaze to Word Mapping (GWM) tool, COM interface description, Version 1.2.3.60. Technical report, University of Tampere.