Literature Survey on Eye-Tracking

Nivvedan S (100050084)

BTP 2013-2014

1 Introduction

Eye-Tracking is the process of measuring eye-positions and eye-movements of a subject. The devices used in carrying out such experiments are known as eye-trackers. Eye-trackers can broadly be classified into:

1. Head-mounted eye-trackers
2. Remote eye-trackers

In head-mounted eye-trackers, sensors are embedded into the glasses worn by subjects while in the Remote eye-trackers the screen is embedded with sensors for recording eye-movements. A headmounted eye-tracker by Eye-Tracker manufacturer `Tobii` is shown in Figure 1[1].The rate at which the eye-tracker captures eye-movements is known as the Sampling Frequency F. Higher the sampling frequency, higher is the number of gaze points recorded by the device in unit time. In linguistic research, remote eye-trackers are normally used since they are more suitable for reading/writing experiments. Modern eye-tracker manufacturers ship *Experiment Design Software* through which eye-tracking experiments can be configured, replayed and analysed. For example, `Tobii` ships a software called `Tobii Studio` along with the device. One can also utilize the APIs provided by the manufacturers to develop special purpose eye-tracking applications.

[1]Image from Manufacturer's website - `http://www.tobii.com/`

Figure 1: A head-mounted Eye-Tracker from `Tobii`

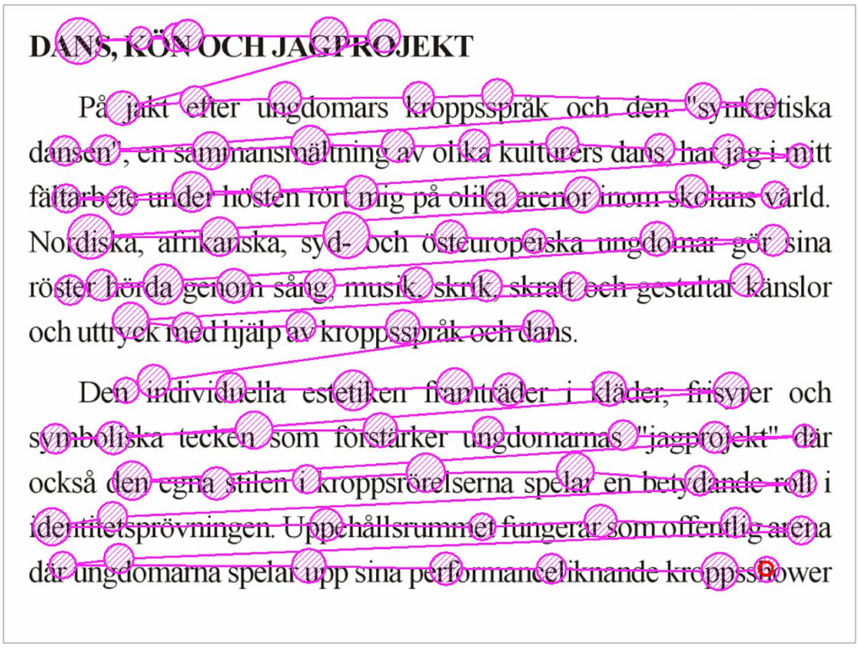


Figure 2: An Illustration of of fixations and saccades over text. The circles represent the fixations and the lines represent the saccades

2 Some Eye-Tracking Terms

Some commonly used eye-tracking terms are defined here.

Gaze

A momentary eye-fixation at a particular point / word. A gaze has no duration associated with it.

Fixation

A prolonged eye-fixation at a particular point / word. Consecutive gazes at proximal locations are aggregated into a fixation. A fixation also has a duration associated with it.

Saccade

Rapid movement of the eye between two fixations. This is better illustrated with Figure 2[2].

Scanpath

A sequence of consecutive eye fixations.

[2] Image from Wikimedia Commons (`https://commons.wikimedia.org/wiki/File:Reading_Fixations_Saccades.jpg`)

3 Eye-Tracking and its applications

3.1 Early Eye-Tracking studies

Eye-tracking experiments in the context of reading studies have an early history. One of the earliest eye-trackers was designed by Edmund Huey (Huey, 1908) which just consisted of a contact lens like device with a hole for the pupil. However, it seems that significant observations on reading behaviour were made even before Huey's early eye-tracker, just by direct observation of eye-movements. Huey, in his book mentions that in 1879, the French ophthalmologist Louis Émile Javalmade makes the profound observation that eye-movements are not as smooth as previously thought, but consists of discreet fixations (short stops) and saccades (rapid eye movements).

Yarbus (Yarbus, 1967), in his book Eye Movements and Vision, wrote about the relationship between fixation and interest in the context reading which opened up eye-tracking studies to psychological and business studies. This is followed by an intense spell of eye-tracking related research in reading studies by psychologists. Rayner (1978) presents a good summary of the eye-tracking psychological research in this period.

3.2 The Just & Carpenter Theory of Reading

The next big milestone in eye-tracking studies is the formulation of the Just and Carpenter reading model and the strong eye-mind hypothesis, in their notable work *A theory of reading: from eye fixation to comprehension* (Just and Carpenter, 1980)

Just and Carpenter developed a theory of reading to account for the gaze-fixations of college students reading scientific passages. Their model hypothesises that *"readers make longer pauses at points where processing loads are greater"* and that *"greater loads occur while readers are accessing infrequent words, integrating information from important clauses, and making inferences at the ends of sentences."*

Their reading model is based on certain observations made by researchers regarding reading behaviour. Specifically, a few of them have been mentioned here below.

It a misconception that readers do not fixate every word but only some small fraction of the text. This has been demonstrated by the data collected by Just and Carpenter and also by experiments performed previously. (Buswell, 1937, chap. 4; Dearborn, 1906, chap. 4; Judd and Buswell, 1922 chap. 2).

McConkie and Rayner reported (1975; Rayner, 1978) that readers generally cannot determine the meaning of a word that is in peripheral vision.

The Just and Carpenter reading model proposes that the gaze duration reflects the time to execute comprehension processes, the longer fixations are attributed to longer processing caused by the word's infrequency and its thematic importance. The link behind eye fixation data and the theory rests on two crucial assumptions:

1. **Immediacy assumption**:
   A reader tries to interpret each content word of a text as it is encountered, even at the expense of making guesses that sometimes turn out to be wrong.

2. **The Strong eye-mind hypothesis**:
   This hypothesis is a very significant contribution of Just and Carpenter and a lot of research following Just and Carpenter would be based on this. As stated by Just and Carpenter, the strong eye-mind hypothesis is as follows:

   The eye remains fixated on a word as long as the word is being processed.

   It can also be stated in an equivalent form as:

   There is no appreciable lag between what is fixated and what is processed.

3.3 The Reading Model

The reading and comprehension process involves the following sub-processes as described by Just and Carpenter:

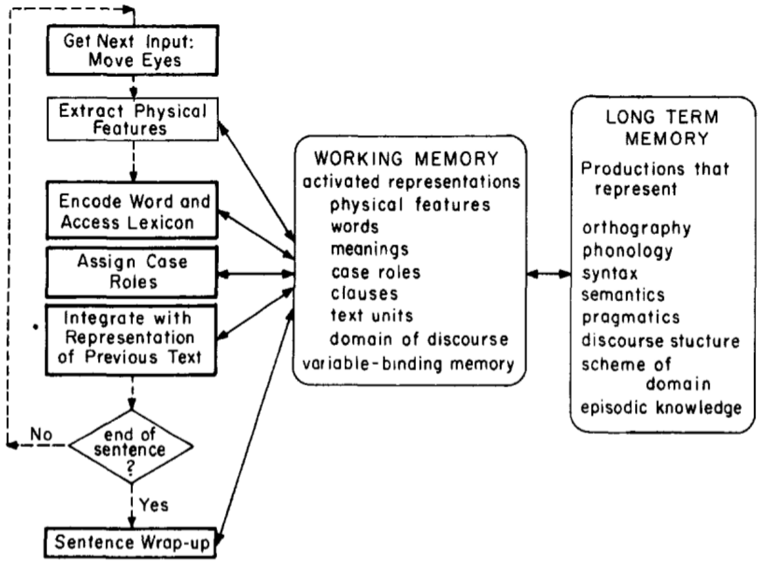
- **Get Next Input** - This is the first stage of a cycle that finds information, encodes it, and processes it.
- **Word Encoding and Lexical Access** - The reading process involves encoding a word into an internal semantic format.
- **Assigning Case Roles** - Comprehension involves determining the relations among words, the relations among clauses, and the relations among whole units of text.
- **Inter-clause Integration** - As each new clause or sentence is encountered, it must be integrated with the previous information acquired from the text or with the knowledge retrieved from the reader's long-term memory.
- **Sentence Wrap-Up** - This special computational episode occurs when a reader reaches the end of a sentence. The processes that occur during sentence wrap-up involve a search for referents that have not been assigned, the construction of inter-clause relations, and an attempt to handle any inconsistencies that could not be resolved within the sentence.

Figure 3 is a schematic diagram of the major processes and structures involved in reading comprehension, from Just and Carpenter, 1980 Figure 1.

3.4 Eye-tracking in NLP

The field of Natural Language processing has seen a huge surge in eye-tracking applications in recent years. On this note, the work done on eye-tracking in the context of Natural Language Processing at IIT Bombay will be discussed in

Figure 3: A schematic diagram of the major processes and structures in reading comprehension [Just and Carpenter, 1980, Figure 1] (Solid lines denote data-flow paths, and dashed lines indicate canonical flow of control)



detail in the subsequent chapter.

Mishra et al. (2013) in their preliminary work on translation difficulty used only fixation duration and saccadic duration as difficulty indicators to label the training data. However, recent literature talks about the information conveyed by scanpaths about sentence processing. This is currently being explored in our on-going work and we shall discuss scanpaths briefly here.

von der Malsburg et al. (2012) showed that scanpaths analysis could be a valuable resource in analysing sentence processing and also discuss freely available, open-source scanpath analysis methods. In traditional eye-tracking analysis techniques the eye-regression movement isn't accounted for properly. However, it has been shown y von der Malsburg et al. (2012) that in complicated sentences, eye-regression movement is very common. In fact, it was demonstrated that for garden-path sentences, around 50% of all saccadic movements are backwards and failure to account for them could result in loss of valuable information.

4 Eye-Tracking Work at IIT Bombay

Over the past couple of years, Eye-tracking techniques have been of considerable interest to the NLP research groups at IIT Bombay. In this chapter, we shall discuss two published works from IIT Bombay that have used Eye-tracking techniques extensively.

The First paper by Joshi et al. (2013) uses eye-tracking to form a hypothesis regarding the cognitive sub-processes involved in Word Sense Disambiguation (WSD). The second paper is the precursor to this report, work by Mishra et al. (2013) on automatically predicting sentence translation difficulty.

4.1 More than meets the eye: Study of Human Cognition in Sense Annotation

Joshi et al. (2013) highlight that the current Word Sense Disambiguation (WSD) systems are fundamentally Weak AI systems. However, according to the classical definition, a strong AI based WSD system should perform the task of sense disambiguation in the same manner and with similar accuracy as human beings. In order to build such a system, it is important to understand the cognitive sub-processes involved in WSD.

An attempt is made to answer two major questions in their work:

- What are the cognitive sub-processes associated with the human sense annotation task?
- Which classes of words are more difficult to disambiguate and why?

Eye-tracking experiments were performed on a generic Hindi news corpus with the help of three skilled lexicographers and three unskilled lexicographers, and fixations and saccades were then measured.

Based on the analysis of the fixations, saccades and scanpaths, the following hypothesis is proposed regarding the cognitive sub-process in WSD.

1. When a lexicographer sees a word, he/she makes a hypothesis about the domain and consequently about the correct sense of the word, mentally. We denote the time required for this phase as T_{hypo}.
2. Next, the lexicographer searches for clues to support this hypothesis and in some cases to eliminate false hypotheses, when the word is polysemous. We denote the time required for this activity as T_{clue}.
3. The clue words aid the lexicographer to decide which one of the initial hypotheses was true. To narrow down the candidate synsets, the lexicographers use synonyms of the words in a synset to check if the sentence retains its meaning.

The following hypothesis is finally proposed:

$$T_{total} = T_{hypo} + T_{clue} + T_{gloss} \tag{1}$$

where
T_{hypo} is the Time for hypothesis building.
T_{clue} is the Clue word searching time.
T_{gloss} is the Gloss Matching time and winner sense selection time.

4.2 Automatically Predicting Sentence Translation Difficulty

Mishra et al. (2013) introduce Translation Difficulty Index (TDI), a measure of difficulty in text translation. The key point with the development of TDI is that, instead of relying on the direct input of the translators for a measure of translation difficulty, it relies on cognitive evidences from eye tracking. TDI is measured as the sum of fixation (gaze) and saccade (rapid eye movement) times of the eye.

The paper considers three parameters that could contribute to translation difficulty:

1. Sentence Length
2. Lexical Properties (Degree of Polysemy)
3. Syntactic Properties (Structural Complexity)

and establishes a strong co-relation of these parameters with empirical data.

TDI of a sentence is taken to be the "*time for which translation related processing is carried out by the brain*" which is which is called the *Translation Processing Time* T_p.

$$T_p = T_{p,comp} + T_{p,gen} \tag{2}$$

Where $T_{p,comp}$ and $T_{p,gen}$ are the processing times for source text comprehension and target text generation respectively.

The empirical TDI, is then computed by normalizing T_p with the sentence length L.

$$TDI = \frac{T_p}{L} \tag{3}$$

The paper claims that measuring T_p is a difficult task as translators often switch between thinking and writing activities, Hence, Eye-tracking is used to do the job.

Since T_p is the sum of fixation and saccadic duration, T_{comp} and T_{gen} can further be broken down as:

$$T_p = \sum_{f \in F_s} dur(f) + \sum_{s \in S_s} dur(s) + \sum_{f \in F_t} dur(f) + \sum_{S \in S_t} dur(s) \tag{4}$$

Here, F_s and S_s correspond to sets of fixations and saccades for source sentence and F_t and S_t correspond to those for the target sentence respectively. *dur* is a function returning the duration of fixations and saccades.

Once the duration is obtained, a support vector regressor is trained with the data to predict the TDI. The SVR performs best with a quadratic kernel, giving 12.88% mean square error with a 10-fold evaluation.

References

G Buswell. How adults read." supplementary educational monographs number 45. chicago: University of chicago press. In *In Sticht and Armstrong*, pages 43–50, 1937.

W Dearborn. The psychology of reading. columbia university contributuions to philosophy and, 1906.

Edmund Burke Huey. The psychology and pedagogy of reading, 1968.

Salil Joshi, Diptesh Kanojia, and Pushpak Bhattacharyya. More than meets the eye: Study of human cognition in sense annotation. pages 733–738, 2013.

C H Judd and G T Buswell. Silent reading: A study of the various types, 1922.

M A Just and P A Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological Review*, (87), 1980.

Abhijit Mishra, Pushpak Bhattacharyya, Micheal Carl, and CRITT, I. Automatically predicting sentence translation difficulty. In *Proceedings of ACL*, 2013.

K Rayner. Eye movements in reading and information processing. *Psychological Bulletin*, (85):618–660, 1978.

Titus von der Malsburg, Shravan Vasishth, and Reinhold Kliegl. Scanpaths in reading are informative about sentence processing. 2012.

A L Yarbus. Eye movements and vision, 1967.