Grounding Translation Tools in Translator's Activity Data

Michael Carl
Department of International Language Studies and Computational Linguistics
Center for Research and Innovation in Translation and Translation Technology
Copenhagen Business School, 2000 Frederiksberg, Denmark

Abstract

This paper presents a technology and a representation for gathering and analysing User Activity Data (UAD) from human translation sessions. We discuss recent advances in the field of translation process research and investigate how insights from this branch of research could be instrumentalised for the design translation tools. New technologies and novel ways of using existing technology could emerge with enhanced knowledge about translator's behaviour and a tight integration of human and machine translation models.

1 Introduction

While traditional human translator education has not been much concerned with novel technological developments and requirements for new sets of translator skills, translation research has increasingly focused on investigating translation processes from a cognitive perspective In recent years.

A change in translation process research can be observed from the earlier study of artificially elicited user data as in think-aloud experiments (Gerloff, 1986; Krings, 1986; Lörscher, 1991) to the more recent study of User Activity Data (UAD) (Carl et al., 2008; Carl and Jakobsen, 2009) of eye movements and keystrokes. By the 1990s most texts and most translations were typed on computer keyboards, and software was developed to log the process by which keystrokes were made in time, ScriptLog (Ahlsén and Strömqvist, 1999), Translog (Jakobsen, 1999). With these tools, a complete log could be created of all the keystrokes made in producing a text, including typos, pauses, deletions, changes, mouse clicks, cursor movements, etc.

One disadvantage about keystrokes is that they are made at the tail-end of the translation or (post)editing process: First there is reading and construction of source text meaning. Then there is mapping of this meaning onto a representation in the target language, and then there is typing of that new representation. What is reflected in the typing activity is the discharge of a segment of information stored in working memory.

Eye movements, in contrast, give a detailed picture of the complex processing involved in constructing meaning from a string of verbal symbols. Fundamentally, reading progresses from left to right (with left-to-right writing systems) along one line at a time and from the end of a line to the beginning of the next line down, but reading is by no means a smooth succession of fixations strung together by forward-moving saccades. Whenever meaning construction fails temporarily, a regressive saccade moves the eyes back to a previous part of the text for reinspection. Fixations differ greatly both with respect to their duration in time and with respect to the number of times one and the same language item may be fixated (Radach et al., 2004; Rayner and Pollatsek, 1989).

The relationship between what the eyes are doing at any given moment in time and what the mind is processing is not as straightforward as was originally assumed by Just and Carpenter (1980). Sometimes the mind is ahead of the

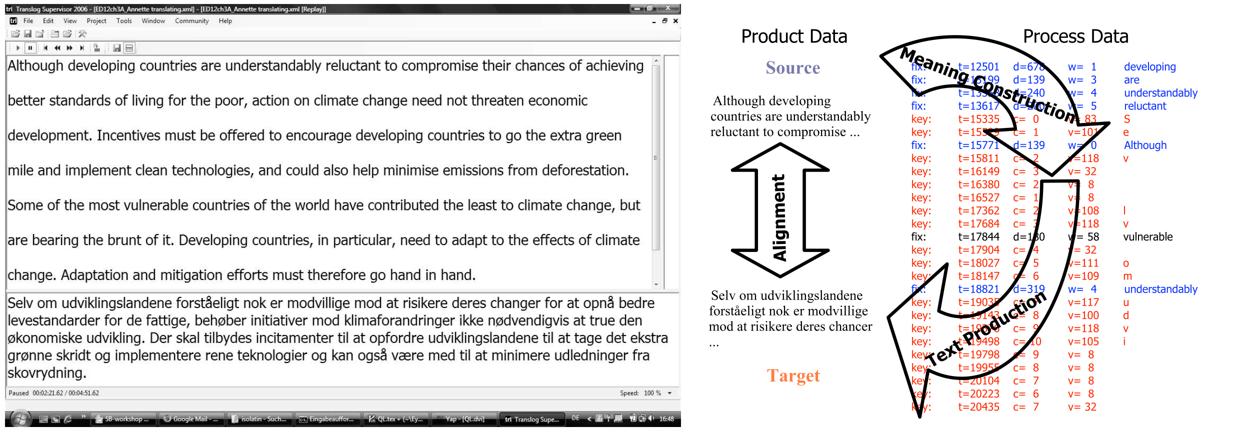

Figure 1: The left side plots a Translog screen shot of a translation session with the source text in the upper window and the produced translation in the lower window. The right side represents the structure of the collected UAD. UAD consists of the textual product data (as visualised in the editor) and the process data, i.e. gaze fixations and keystoke actions.

eyes and is already processing information represented by a word the eyes have not yet fixated. Sometimes the eyes move ahead so fast that the mind lags behind and has to catch up. Such temporal misalignment may cause an earlier or a later word to be fixated longer even if the processing concerned a neighbouring item. Likewise there are at least three different ways in which the eyes may respond to processing difficulty: they may fixate an item longer, they may move on (and fixate a subsequent word while they wait for the mind to catch up), or they may execute a regressive saccade and refixate words already read. Liversedge and Findlay (2000) have proposed to deal with such complexities by means of hybrid eye movement parameters which aggregate fixation patterns across several words.

Optimal human translation (or post-editing) would involve that a constant supply of processed ST meaning and TT mapping was fed into working memory at a rate that would allow the translator to type continuously at maximum speed. However, since this situation rarely obtains for intervals longer than about half a minute (Jakobsen, 2005), text production keystrokes tend to be clearly segmented into units reflecting the chunks of meaning that were processed either immediately before the keystrokes were made or starting before but overlapping to some extent with the period of typing.

It is, however, largely unclear how this process is organised in detail, what it is exactly that makes a text difficult to translate, how translation difficulties can be detected and how a translator could be helped in a best way. Doherty and O'Brien (2009), for instance, present a study which indicates that the quality of (machine translated) sentences can be detected by the number of fixations and the overall sentence reading time. Awkward translations would need more time to be read per character than good ones. These findings support *that* eye-tracking might be instrumentalised to support translators during post-editing MT output. However, in order to know *how* this could be done, we need to push the analysis a step further and look deeper into the details of human translation processes.

Reading disfluencies might be due, for instance, to unknown or unusual words, awkward, confusing or complicated sentences. Difficulties in text construction are visible in keyboard patterns such as pauses, deletion, and correction patterns, lexical substitution, movement of textual elements, etc.

This paper looks at the interplay of gaze and keyboard patterns during human translation activities and investigates their temporal properties. The goal is to detect typical patterns of

1. Source text: `<word str="Although" cur="0" pos="ADV"  top="76" btm="110" ... />`
   includes word information, information of word location on the screen and in the text.
2. Target text: `<word str="Selv" cur="0" top="511" btm="544" lft="21" rgt="69"  />`
   same information than source text representation.
3. Alignment source-target text: `<align src="0" tgt="0"/>`
   indicates which word(s) are translations of each other.
4. Eye gaze data: `<eye time="140" lx="748"  ly="122" lp="3.853" ... />`
   includes pixel location for the left eye (if available also for the right eye) and pupil dilation.
5. Fixation information: `<fix time="12501" dur="678" cur="9" ... />`
   starting time and duration of fixations as well as mapping onto fixated word.
6. Keyboard activity: `<key time="15346" val="83" cur="1" str="S" ... />`
   timestamp of keyboard action, cursor position and key value.

Figure 2: Representation of UAD in six co-indexed files

fluent and of disfluent reading and writing and to link these patterns to properties of the source and the produced target text. We believe that a better understanding of the underlying processes will give us also targeted means to develop enhanced tools for translators.

We look first into peculiarities of UAD and outline a suitable XML format. In section 3 we describe a query language for this data and show how patterns of UAD can be retrieved and correlated. In section 4 we provide an analysis of a human translation session. Section 5 discusses the results and gives a prospective about translation aides grounded in an understanding of UAD.

2 The Structure of UAD

As shown in figure 1, UAD relates spatial i.e. textual product data with temporal process data. Fixations are linked to word positions on the screen, and temporal sequences of keystrokes result in static text translations. Figure 1 suggests that meaning construction of a source text is preliminary to the production of the target language translation. To investigate this process and to discover the dependencies between process and product data, we suggest a linked and queryable six-dimensional data representation as shown in figure 2.

The *product data* consists of three resources, the source text, the target text (i.e. the translation) and a link between entities of both texts. The location for each word in the source and target texts is identified by its position on the screen and its position in the text. The screen position of the word is indicated with the top-left and bottom-right pixel position while cursor positions give the character offset from the beginning of the text. Pixel positions information is essential for gaze-to-word mapping, while cursor positions are used as an index of the word.

The *process data* consists equally of three resources, the gaze-sample points, fixations and keystroke information. Gaze sample points consist of a left- (and/or right-) eye position as well as pupil dilation at a particular time. With our current eye-tracker (TOBII 1750) the time interval between successive gaze-samples points is 20ms. A fixation groups together a number of near-distance eye-gaze samples which represent a time segment in which a word is looked at. Fixations have a starting time, a duration and a cursor position. The cursor position refers to an index in the source (or target) text[1]. Keyboard activities consist of a key value at a cursor position in the target window and a time stamp.

3 A Query Language for UAD

With this representation we are able to relate dimensions in the UAD via cursor positions (in the

[1] We are currently only able to collect fixations on the source text.

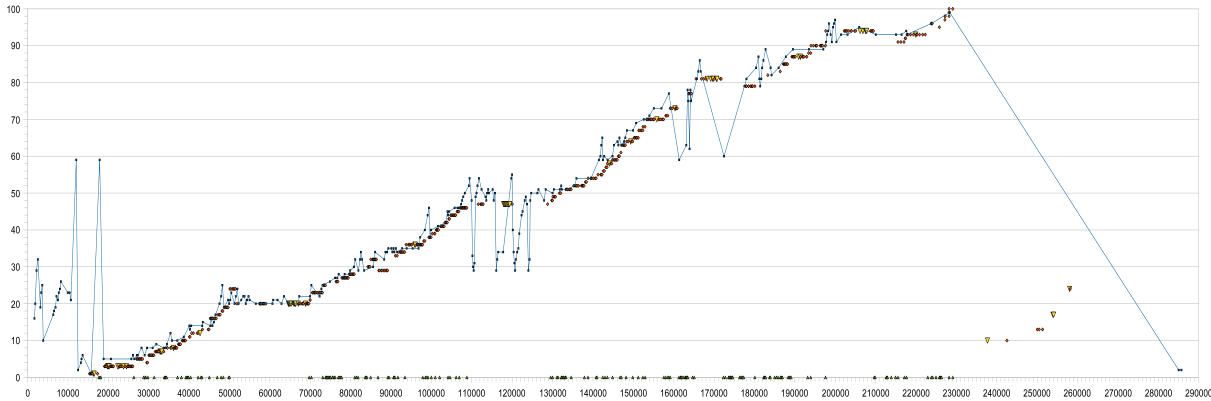

Figure 3: The graph depicts the translation progression of the experimental setting in figure 1: the vertical axis enumerates the words of the source text, the horizontal axis represents the translation time. The dots show how the target text evolve in time.

product data) or via timestamps in the process data. In addition, a query language is needed to formulate, retrieve and relate patterns in the product and product data so as to reveal correlations between them.

Patterns of fixation would give insight into how a sentence is parsed, and where difficulties in meaning construction occur. Patterns of keyboard activities would show how text production is structured, and where problems in target language generation occur. The query language would allow to correlate and to link those patterns.

UAD is represented as a 6-tuple `{S,T,A,E,F,K}` for Source and Target text, Alignment, Eye, Fixation and Keyboard data respectively. The query language allows to describe patterns which are mapped onto the data dimensions. Nodes in the UAD are addressed with sets of "`attribute.operator.value`", and variables (starting with "`$`") of the language are instantiated with the values of the currently matching part.

A pattern rule, as in table 1, may consists of several successive patterns which describe different dimensions of the UAD. Variables (e.g. `$A1`, `$A2`) which are instantiated in one of the dimensions (e.g. `cur.V.$A1`, in line 1) may serve as a constraint in another dimension (e.g. `src.eq.$A1`, in line 2). In this way several dimensions of the UAD can be linked and correlated. Basically, each pattern matches successively every node in the data and instantiates the variables with the current values. If a pattern successfully matches, the rule switches to the next line until either a pattern fails or the end of the data was reached. The rule will then backtrack and search for the next pattern in the previous line.

```
1  S>[str.eq.developing,cur.V.$A1]
2  A>[src.eq.$A1,tgt.V.$A2]
3  T>[cur.eq.$A2,str.V.$V1]
4  !:InitKeyRange('Key', $A2-1, $A2+length($V1));
5  K<kp[cur.val.Key]
6  !:PrintPattern('kp:time,type,cur,str');
```

Table 1: The rule prints all keyboard activities related to the translation of "developing".

The rule in table 1 retrieves all keystrokes that are related to the production of the translation for "developing". The patten in line 1 iteratively matches all instances of "developing" in the source text and instantiates the variable `$A1` with the cursor positions. The pattern in line 2 retrieves the cursor position of the translation via the alignment data (`A`) and the pattern in line 3 retrieves the word form (e.g. "`udviklingslandene`") of the translation from the target text (`T`) and stores it in the variable `$V1`.

The keystroke pattern (K) in line 5 looks from backwards into the keyboard data and marks all keystroke nodes with the marker `kp` which contribute to the production of "`udviklingslandene`". A special operator `val` is used to keep track of the beginning and end cursor positions in the target text segment. `InitKeyRange` initialises these values with the beginning position `$A1-1` and the end position `$A1+length($V1)` of the traced translation.

In this way we can correlate different UAD dimensions. Figure 3 represents the retrieved correlation in a compact form, which will be discussed in the next section.

4 An Example Analysis

The graph in figure 3 represents a translation session of the 100 word text in figure 1 from English into Danish. The vertical axis in figure 3 enumerates the words of the source text (from 1 to 100) while the horizontal axis represents ca. 5 minutes (almost 290.000 msec.) in which the Danish translation was generated. The graphs plot fixations on the source text[2] (solid line) and keyboard activities (small dots) during the 5 minutes translation time. All dots on one horizontal line represent keystrokes for the translation of the source word represented by that line. The extraction of the keyboard data was achieved with a rule as discussed in table 1 where all keyboard activities were extracted for each source word.

In figure 3, the translator starts with an "orientation phase" of ca. 15 seconds. Some words in the source text are grasped, but no keystrokes are produced. The translator then starts typing while the eyes are in most cases not more than 2 or 3 words in the source text ahead of the currently typed translation. In the beginning, the translation evolves linearly, where the target words occur in the same order as the source text. Around word number 20 after ca. 50 seconds, a syntactic reordering seems necessary, and some time is spent on the translation for Source word 20.

Two larger orientations with translation pauses occur roughly between seconds 110-130 and 160-180. There is also a short post-editing phase from seconds 240-260 in which some words in the beginning of the text are corrected.

Translation pauses represent sequences of UAD where no keyboard actions take place (Jakobsen, 2005) and which only consist of fixations. Translation pauses represent the span of time in which a new chunk of meaning is constructed from the source text. The fixation pattern in figure 4 shows in more detail the context of a longer translation pause in the timespan from seconds 80 to 150. It is extracted from figure 3. The (green) triangular dots on the horizontal 0-axis represent keyboard activities which could not be attributed to the production of any target word.

In figure 4, the translator produces fluent target text, with only little look-ahead in the source text, until around source word 46 where long regressive saccades occur. The sequence of characters produced between timestamps 110.000 and 115.000 is again deleted just before timestamp 120.000. This translation pause lasts approximately 20 seconds, from sec. 110 to 130. After this linear translation production continues.

The fixated sentence in figure 5 contains a conjunction of two main clauses, where the second main clause has an elliptical subject. The translation pause starts with a few regressions from the word "extra" to the beginning subject "Incentives" of the first clause, then refixates the verb in the first clause "offered", and from there goes back to the position from where the repression started, on the verb "could" in the second clause. From there a number of progressions lead through the main clause up to the finite verb of the second main clause. Here the reader obviously notices the missing subject and verifies whether the subject of the first main clause also suits the second main clause. To do so, the reader goes back via the conjunction "and" to the first subject "Incentives" and from there back to the position where she previously was. After dwelling for some time on the following

[2] As said earlier, fixations on the emerging target text are not recorded. During longer spans without recorded fixations, e.g. between seconds 230-290 and around 180, the translator is likely to look into the target text.

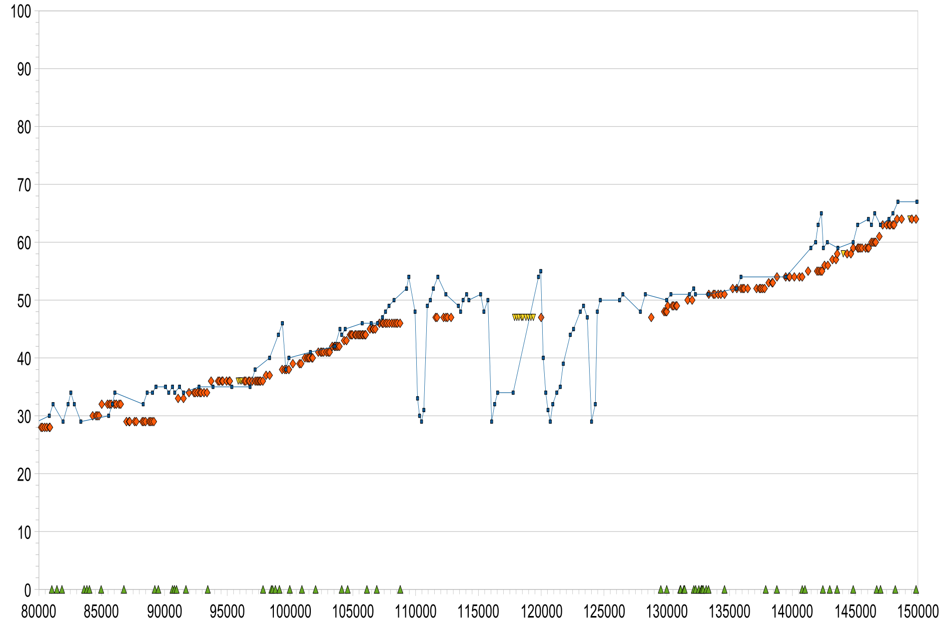

Figure 4: The graph is an extract from figure 3 between seconds 80 to 150. It shows the context of a fixation pattern which is related to the translation problems of the source language text around word number 46.

words (“could”, “help”, “minimise”), a decision was taken to start typing the translation.

In addition to the graphical representation, figure 5 also gives fixation durations, pupil dilation and the number of skipped characters between successive fixations (trans). It is interesting to see that saccades across more than 100 characters take place with high precision. Surprisingly, no fixation was registered on “green mile” and “technologies”, maybe due to imprecision of our gaze-to-word mapping software, especially at the outer edges of the screen. However, the close-by words “the” and “clean” have the longest fixation times with 458 and 459ms, and “green mile” and “technologies” may well have been within parafoveal or peripheral scope when those words were fixated.

5 Discussion

The paper presents a method to systematically investigate and quantify UAD. It discusses an example that shows how translation problems can be detected and qualified. There are many open issues that emerge from this investigation relating to the question of how automated assistance could help the human translation process.

Looking at the progression graph in figure 3: At what moment would the mechanical help be most welcome? Would a translator be better supported during the “linear” translation production or during the translation pauses?

Will it be possible to figure out whether a translation pause is due to unknown terminology (in which case the presentation of a term translation might be an appropriate solution), or for instance, as in figure 4, whether it is due to a more complicated understanding (and translating) problems? Will it also be possible to present appropriate help in the latter case, or would the automated help be distracting?

How should this help be presented? Should a translator be provided with translation completion (à la TransType2) or should he be offered a selection of (phrase) translations which s/he could compose into a translation (e.g. via key shortcuts, as Koehn and Haddow (2009) suggest). How much information should be there, how should the information be structured and how can information overkill be avoided?

Instead of (or in addition to) contemplating

Figure 5: The figure shows the fixation pattern from the translation progression graph in figure 3 between seconds 120-128.
Above: a graphical representation of the fixation pattern together with the sequence of fixated source text. Bold lines represent progressions and light lines regressions.
Below: Process data of the same fixation pattern shows time of fixation, average pupil dilation, fixation duration, cursor position and word number as well as number of characters between successive fixations.

Time	Dilat.	Duration	Cursor	#word	trans	Word
120162	3.48	139	277	39	0	extra
120361	3.47	120	236	33	-41	encourage
120541	3.50	139	222	30	-14	be
120720	3.51	179	206	28	-16	Incentives
120919	3.35	279	225	31	19	offered
121238	3.16	220	236	33	11	encourage
121518	3.15	199	246	34	10	developing
121757	3.15	458	273	38	27	the
122315	3.15	239	298	43	25	implement
122574	3.19	459	308	44	10	clean
123113	3.16	239	332	47	24	could
123372	3.12	279	338	48	6	also
123671	3.13	259	328	46	-10	and
124010	3.08	239	206	28	-112	Incentives
124309	3.11	139	225	31	19	offered
124468	3.12	220	332	47	107	could
124708	3.21	318	343	49	11	help
126243	3.26	239	343	49	0	help
126502	3.26	399	348	50	5	minimise
127898	3.27	378	332	47	-16	could
128296	3.27	319	348	50	16	minimise

these questions independently, we might also aim at a stronger, more fundamental integration of the translator and the translation system based on an isomorphism of the translation processes: We have almost identical scenarios in the study of translator's activity data and in statistical machine translation: in both cases we "generate a story of how an English string e gets to be a foreign string f"[3]. In this paper, we have presented a procedure that traces how a translation e emerges from keystrokes and we can pin-point the gaze activities that precede the translation generation, reflecting the process of understanding (the relevant parts of) the source text f. Unlike in SMT, where the "choices in the story are decided by reference to parameters $p(f|e)$", we use eye-tracker and keyboard logging devices to detect translation problems (i.e. choice cases) in the gaze and text-production rhythm, which, given enough data, might equally lead to probabilistic models.

While the presented method makes it possible to detect, categorise and quantify as yet unobserved properties in human translation processes, a real-time integration into new translation tools still awaits many problems to be solved. To the extent we better understand the UAD, we will also understand how human and machine translation processes may be mapped and better integrated.

References

E. Ahlsén and S. Strömqvist. 1999. ScriptLog. A tool for logging the writing process and its possible diagnostic use. In *(Loncke et al., 1999)*, pages 144–149.

Michael Carl and Arnt Lykke Jakobsen. 2009. Investigating Translators Activity Data. *International Journal of Speech Technology*, page forthcoming.

Michael Carl, Arnt Lykke Jakobsen, and Kristian T.H. Jensen. 2008. Studying Human Translator Behavior with User-Activity Data. In *Natural Language Processing and Cognitive Science (NLPCS 2008)*, Barcelona, Spain.

Helle V. Dam, Jan Engberg, and Heidrun Gerzymisch-Arbogast, editors. 2005. *Knowledge Systems and Translation*, volume 7 of *Text, Translation, and Computational Processing*. Mouton de Gruyter, Berlin, New York.

Stephen Doherty and Sharon O'Brien. 2009. Can MT Output Be Evaluated Through Eye Tracking? In *MT Summit*, page This conference.

P. Gerloff. 1986. Second Language Learners Reports on the Interpretive Process: Talk-aloud Protocols of Translation. In *(House and Blum-Kulka, 1986)*, pages 243–262.

Gyde Hansen, editor. 1999. *Probing the process in translation: methods and results*, volume 24 of *Copenhagen Studies in Language*. Copenhagen: Samfundslitteratur.

Juliane House and Shoshana Blum-Kulka, editors. 1986. *Interlingual and Intercultural Communication*. Discourse and Cognition in Translation and Second Language Acquisition Studies. Gunter Narr, Tübingen.

Arnt Lykke Jakobsen. 1999. Logging target text production with Translog. In *(Hansen, 1999)*, pages 9–20.

A. L. Jakobsen. 2005. Investigating expert translators processing knowledge. In *(Dam et al., 2005)*, pages 173–189.

M.A. Just and P.A. Carpenter. 1980. A theory of reading from eye movements to comprehension. *Psychol. Rev.*, 87:329–354.

Philipp Koehn and Barry Haddow. 2009. Interactive Assistance to Human Translators using Statistical Machine Translation Methods. In *MT Summit*, page This conference.

H. Krings. 1986. *Was in den Köpfen von Übersetzern vorgeht*. Gunter Narr, Tübingen.

S.P Liversedge and J.M. Findlay. 2000. Saccadic eye movements and cognition. *TICS*, 1(4):6–14.

F. Loncke, J. Clibbens, H. Arvidson, and L. Lloyd, editors. 1999. *Augmentative and Alternative Communication*. New directions in research and practice. Lawrence Erlbaum Associates, Mahwa, New Jersey.

W. Lörscher. 1991. *Translation Performance, Translation Process and Translation Strategies. A Psycholinguistic Investigation*. Gunter Narr, Tübingen.

R. Radach, A. Kennedy, and K. Rayner. 2004. *Eye movements and information processing during reading*. Psychology Press, East Sussex.

K. Rayner and A. Pollatsek. 1989. *The Psychology of Reading*. Lawrence Erlbaum Associates, Hillsdale, New Jersey Hove, UK.

[3] http://people.csail.mit.edu/koehn/publications/tutorial2003.pdf