Objectives for a Query Language for User-activity Data

Michael Carl and Arnt Lykke Jakobsen

Copenhagen Business School
Department of International Language Studies & Computational Linguistics
2000 Frederiksberg, Denmark

Abstract. One of the aims of the Eye-to-IT project (FP6 IST 517590) is to integrate keyboard logging and eye-tracking data to study and anticipate the behaviour of human translators. This so-called User-Activity Data (UAD) would make it possible to empirically ground cognitive models and to validate hypotheses of human processing concepts in the data. In order to thoroughly ground a cognitive model of the user in empirical observation, two conditions must be met as a minimum. All UAD data must be fully synchronised so that data relate to a common construct. Secondly, data must be represented in a queryable form so that large volumes of data can be analysed electronically.

Two programs have evolved in the Eye-to-IT project: TRANSLOG is designed to register and replay keyboard logging data, while GWM is a tool to record and replay eye-movement data. This paper reports on an attempt to synchronise and integrate the representations of both software components so that sequences of keyboard and eye-movement data can be retrieved and their interaction studied. The outcome of this effort would be the possibility to correlate eye- and keyboard activities of translators (the user model) with properties of the source and target texts and thus to uncover dependencies in the UAD.

1 Introduction

In a recent paper [1] we presented a number of advantages that could be obtained by studying User Activity Data (UAD) to investigate human translation behaviour. UAD consists of the translator's recorded keystrokes and eye-movement behaviour, and their link to the read source and to the produced target text and to the translator's processing behaviour. While eye movement patterns reflect how (source) text is parsed, how meaning is constructed in working memory [3, 5] and how, in most cases, text production is monitored, keyboard activities reflect the discharge of this chunk of information and the unfolding of the meaning in the target language. One major advantage about working with UAD is that data are not skewed by concurrent think-aloud or by intrusive methods, but the greatest advantage is that UAD opens up the possibility of empirically grounding a cognitive model of translator behaviour based on event coding [4]. Such a model will give us new insight into translational processing, what goes on in the translator's mind [5], and it will make it possible for us to construct help tools that will "know" and sometimes anticipate the user's needs and therefore will be able to offer immediate on-the-fly assistance.

In order to thoroughly ground a cognitive model of the user in empirical observation, two conditions must be met as a minimum. All UAD data must be fully synchronised so that data relate to a common construct [2]. Secondly, data must be represented in a structured form and stored in a database which can be queried. This will make it possible to identify reoccurring patterns of UAD phenomena and to align them with text items or properties (in the source and/or target texts) and associate them with basic processing concepts in the user model. A query language designed for this purpose will allow the researcher to handle very large volumes of empirical data and so to establish connections between UAD and textual phenomena and basic processing assumptions grounded solidly in empirical data from highly naturalistic translation events.

Within the Eye-to-IT project, two different software components are currently used to record keyboard activities and eye-movement patterns: GWM records gaze-movements and computes gaze-to-word mapping hypotheses while TRANSLOG records keystrokes. Both programs use different timers, different text indexing strategies and different formats of representation so that it is difficult to synchronise the information perfectly with millisecond precision. Both programs have a replay mode to review a translation session, which can then be commented on by the translator in a retrospective interview.

A screen shot of the experimental setting is shown in figure 1. The screen shot shows two windows: the upper window contains the English source text; the lower window its partial translation into Danish. User activity data were collected from the typing activities in the lower window and reading activities in both the upper and lower windows[1].
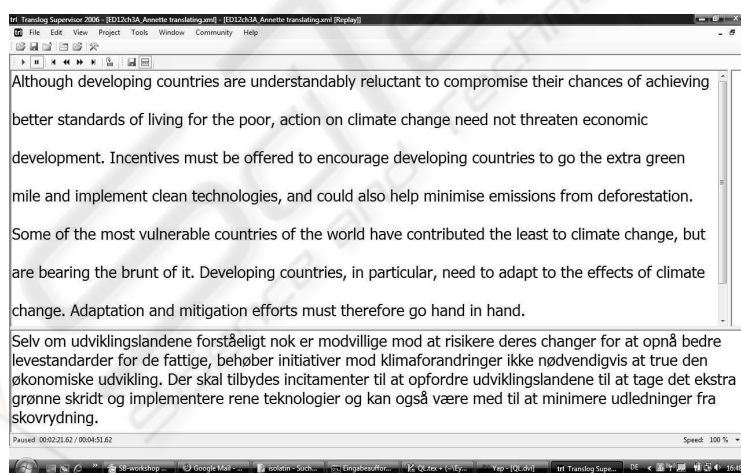

Fig. 1. The figure shows the TRANSLOG replay screen at time 2 mins and 21.62sec. after the beginning of the translation session. The upper window plots the source text to be translated, while the lower window shows the produced target language text (Danish).

---

[1] The GWM version that was used in the experiment from which data have been used only recorded fixations in the upper window.

The present paper reports an attempt to integrate and sequentialise keyboard logging and eye-tracking data from TRANSLOG and GWM so that it becomes queryable. A queryable format which integrates both modalities is a prerequisite for advanced modelling of translators which seeks to empirically quantify hypotheses about meaning construction when reading a source text and subsequently writing a translation of it.

In section 2 the paper outlines a format for User-Activity Data. These data were semi-automatically converted from the output of GWM and TRANSLOG for two translation experiments.

Section 3 discusses details for the design of a query language of UAD. For the design of a query language it is important to know what aspects in the data we want to retrieve, correlate and quantify. We might be interested in finding patterns of process data (e.g. regressions or sequences of very long fixations) which correspond to particular properties of product data (e.g. idioms, compounds or cognates) and/or we might be interested in finding sequences of product data that are linked to certain properties in the process data.

Sections 4, 5 and 6 provide examples of what could be queried and retrieved from a base of UAD. Section 4 gives an example which illustrates the history of keyboard activities in the creation of a word. Section 5 looks at the inner structure of a translation pause and section 6 computes fixation bi-grams.

2 Representation of UAD

Despite their different representations, it was possible to synchronise the output of GWM and TRANSLOG in a semi-automatic process and to convert the representation into an XML format. User activity data is distributed in several files so that it can be linked and pieces of information can be related. According to their nature and modality, UAD is grouped together in five files:

1. source text information is shown in table 1: it includes linguistic information and information about the location of words on the screen .
2. basic eye-tracking information is shown in table 2, including pixel location for the left (and if available also for the right) eye, as well as pupil dilation.
3. fixation information is plotted in table 3: the starting time and duration of fixations as well as their mapping onto fixated words.
4. keyboard activity information is shown in table 4: including cursor position and key value
5. target text information is shown in table 5: like the source text representation, it contains linguistic information and information about the position of the word on the screen and in the text.
6. source-target text alignment data is shown in table 6: for each word it indicates to which words in the target language it is aligned.

The location for each word in the source and target texts is identified by its top-left and bottom-right pixel position as well as their cursor positions. Pixel positions are needed for gaze-to-word mapping, while cursor positions are used as an identifier and index for the word.

Table 1. Representation of source text: each word is annotated with linguistic and location information. The word position of the variable adverb with the value "Although" is from cursor position 0 in a line, the top of which is 76 pixels and the bottom of which is 110 pixels from the top of the screen.

```
<text id="source">
  <word pos="ADV" val="Although" cur="0"  top="76" btm="110" ... />
  <word pos="VVG" val="developing" cur="9"  top="76" btm="110"  ... />
  <word pos="NN2" val="countries" cur="20" top="76" btm="110" ... />
  ...
</text>
```

Table 2. Eye-gaze sample points consist of a left-eye position (lx/ly) as well as pupil dilation at a particular time. Note that the time interval between successive samples is 20ms.

```
<sample>
  <eye time="140" lx="748"  ly="122" lp="3.853" ... />
  <eye time="160" lx="757" ly="121" lp="3.788" ... />
  ...
</sample>
```

Table 3. Fixations have a starting time and a duration. The cursor position refers to an index in the source (or target) text.

```
<fixation>
  <fix time="12501" dur="678" cur="9" ... />
  <fix time="13199" dur="139" cur="30" ... />
  ...
</fixation>
```

Table 4. Keyboard activities consist of a key value on a particular cursor position in the target window and a time stamp.

```
<keyboard>
  <key time="15346" val="83"  cur="1" ... />
  <key time="15540" val="101" cur="2" ... />
  ...
</keyboard>
```

Table 5. Like the source text, the target text is also stored as a mixture of screen location data and linguistic information. The target text can be re-constructed from keyboard activities.

```
<text id="target">
  <word cur="0" top="511" btm="544" lft="21"  rgt="69"  val="Selv"/>
  <word cur="5" top="511" btm="544" lft="77"  rgt="115" val="om"/>
  ...
</text>
```

Table 6. The alignment information are the word initial cursor positions of translations in source and target text. Since the format consists of m to n links between translations, also discontiguous chunk translations can be represented.

```
<text id="target">
  <align src="0" tgt="0"/>
  <align src="9" src="20" tgt="8"/>
  ...
</text>
```

Fixation-to-word mappings are expressed by means of cursor position `cur` in table 3. While fixations group together a number of near-distance eye-gaze samples (cf. table 2), they also represent a time segment in which the fixated word is processed. Patterns of fixation thus give insight into how a sentence is parsed, and maybe where difficulties in meaning construction occur.

The cursor positions `cur` of keyboard activities in table 4 represent locations of characters in the produced target language text of the lower window in figure 1. Since the target text may change dynamically, these cursor positions do not necessarily refer to the cursor position of the final translation as in table 5. But the association with the unique key time value makes their value uniquely identifiable. An example of this will be provided in section 4.

User Activity Data relates spatial data, i.e. textual product data, to temporal process data of the translation activity. Thus, sequences of fixations are connected with patterns of word positions on the screen, and spatial distributions of words on the screen are related to temporal processing patterns[2]. Similarly, successive keyboard activities are causally related to the words which they create, and the positions of words in the target text can be explained by the preceding typing activities. In this way, it becomes possible to study the interplay of fixation and keyboard patterns, to investigate their temporal properties and the way they correlate with the spatial distribution of text.
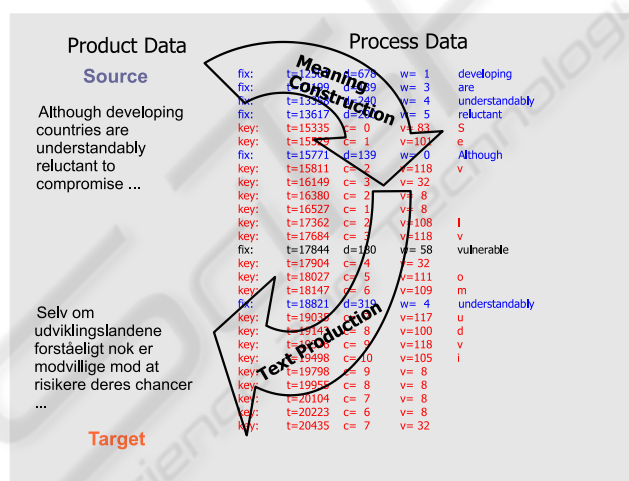

Fig. 2. The figure plots UAD for a small set of data as obtained during a translation session shown in figure 1. The left side shows a segment of the English source text and its translation into Danish. The right side plots a sequence of process data, i.e. gaze fixations and keystoke actions. The sudden and strange fixation on "vulnerable" at time `17844` occurs during a gaze transition between the lower and the upper window.

---

[2] From the data it was only possible to extract fixations on the source text in the upper window in figure 1.

3 Outline for the Design of a Query Language for UAD

With the representation outlined in section 2 we are in a position to retrieve and compare patterns of UAD from sets of data. One goal of this activity is to detect typical patterns of fluent and of disfluent reading and writing and to link these patterns to properties of the source and the target text. Reading disfluencies might be due to unknown or unusual words, awkward, confusing or complicated sentences, while difficulties in text construction are visible in keyboard patterns marked by lengthy pauses. Writing disfluencies may encompass all levels of linguistic description of a word, sentence or inter-sentence level, and may reflect deletion, insertion, correction of typos or lexical substitution, movement of textual elements.

Figure 2 shows the interplay of product and process data. Fixations as in table 3 and keyboard actions from table 4 are represented sequentially, so that temporal relations may be retrieved and studied. As the figure suggests, meaning construction of a source text chunk is preliminary to the production of the target language translation and the dependencies can be observed in the data. A query language would interrogate the database both for product data (source and target text data) and for process data (UAD) and help establish correlations between them and possibly also help associate data with processing concepts in the user model.

The design of a query language for UAD should open the possibility of investigating the data from several points of view. We might be interested in retrieving and comparing any combination of patterns of:

- fixations on particular sequences of texts (e.g. compounds, metaphors, technical terms), asking, for instance: what fixation patterns typically occur on a certain passage of text.
- keyboard actions which lead to a particular passage of the target text, asking, for instance: what are the typing patterns for a certain word or person.
- source texts which satisfy certain fixation patterns. We might be interested, for instance, in investigating:
  - fixation patterns: sequences of non-interrupted fixations (no intervening keyboard activities)
  - progressive fixation patterns: sequences of non-interrupted fixations where for all successive fixations at times t and $t + 1$ the fixated cursor position $cur(t + 1) \geq cur(t)$
  - regressive fixation patterns: sequences of non-interrupted fixations with all fixations at times t and $t + 1$ the cursor positions $cur(t + 1) \leq cur(t)$
- target texts which were constructed with particular keyboard patterns, such as:
  - only appending text (fluent writing)
  - particular deletion patterns
  - modification or re-arrrangement of text
- UAD which occur between the fixation of a SL word and the production of its translation in the TL window.

The remainder gives examples for some of those patterns.

Table 7. Keyboard activities for generating the Danish word "Selv".

Action	Time	Cursor	Value	Char
Key:	15335	0	83	S
Key:	15529	1	101	e
Key:	15811	2	118	v
Key:	16149	3	32	
Key:	16380	2	8	
Key:	16527	1	8	
Key:	17362	2	108	l
Key:	17684	3	118	v

4 Keyboard Patterns

The keyboard activities in table 7 are a subset of the data in figure 2. They plot the typing performed in writing the word "Selv" and were retrieved by querying the process data which corresponds to the cursor positions 1 to 4 of the target text.

The keyboard actions represent a correction of the mis-typed word prefix 'Sev ' where the letter 'l' is inserted after the prefix 'Se'. This is achieved by deleting the sequence 'v ' and replacing it with 'lv'. Note that the value of 'Cursor' in table 7 indicates the position after the keyboard action and 'Value' is the UNICODE of the typed character. The 'Value' 8 signifies a character deletion. Accordingly, the cursor position is decreased after a character is deleted. The average time interval between two successive keyboard actions is around 200-300ms. However, it is almost 800ms for inserting the 'l'.

5 Translation Pauses and Fixation Pattern

Translation pauses are an interesting topic of study, since they represent the span of time in which a new chunk of meaning is constructed from the source text. They represent sequences of UAD which only consist of fixations, and where no keyboard actions take place. The fixation pattern in figure 3 represents a translation pause which was retrieved from the UAD by searching for a non-interrupted sequence of fixations.

A graphical representation of the fixation pattern together with the sequence of fixated source text (the product data that belong to it) is shown in figure 3. The fixated sentence contains a conjunction of two main clauses, where the second main clause has an elliptical subject which it shares with the first clause.

It can be seen that 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 '(must be) 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, quite directly, goes back via the conjunction 'and' to the first subject 'Incentives' and
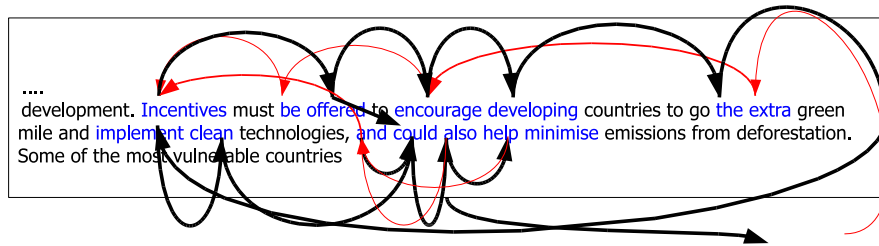
Fig. 3. Above: Structure of a fixation pattern: bold lines represent progressions and light lines regressions.

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

from there back to the position where she previously was. After dwelling for some time on the following words (could, help, minimise), a decision was taken to start typing the translation of the first main clause, since, presumably, an understanding of the second clause was not necessary at that point.

In addition to the graphical representation, figure 3 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.

There is also variation in the pupil dilation during fixations. The average fixation pupil dilation was calculated as a mean over the gaze samples. The highest pupil dilation values were at the beginning of the translation pause, during fixation regression. Dilation decreases after the subject was found and left-to-right parsing of the sentence could take place. Since wider pupil dilation is generally assumed to correspond to larger workload [4, 3], this may indicate that regressions require more cognitive effort than normal 'progressive' reading. This hypothesis is, however, not supported by the second regression in which the subject is sought.

Table 8. Number (#trans) of fixation transitions accumulated from two users over the first 8 words of the source text fragment in figure 1. The columns 'from' and 'to' indicate the word number of fixation departure and landing, 'from word' and 'to word' the respective words.

P	#trans	from#	to#	from word	to word
0.25	1	1	3	developing	are
0.75	3	1	2	developing	countries
0.33	1	2	3	countries	are
0.66	2	2	4	countries	understandably
1.00	2	3	4	are	understandably
1.00	8	4	5	understandably	reluctant
0.23	3	5	4	reluctant	understandably
0.15	2	5	5	reluctant	reluctant
0.38	5	5	7	reluctant	compromise
0.07	1	5	1	reluctant	developing
0.15	2	5	6	reluctant	to
0.5	1	6	5	to	reluctant
0.5	1	6	7	to	compromise
0.25	2	7	5	compromise	reluctant
0.25	2	7	7	compromise	compromise
0.37	3	7	8	compromise	their
0.12	1	7	0	compromise	Although
0.5	1	8	7	their	compromise
0.5	1	8	4	their	understandably

6 Fixation Bi-grams

While the previous examples look at some qualities of individual patterns, a query language should also offer a possibility of correlating and quantifying patterns of UAD across different users and different texts and of generating statistics on the distribution of these patterns. Table 8 shows an attempt to extract fixation bi-grams from two translation experiments. It gives the numbers and relative frequencies of some fixation transitions, as well as the word positions and the actual words fixated.

Collected over a large sample of UAD, fixation bi-grams, or for that matter fixation n-gram models, may indicate the likelyhood of patterns encountered in different kinds of reading as, for instance, reading for understanding, reading for translation, reading technical text, reading web-pages or even disturbed reading of ungrammatical texts. Just like n-gram language models are a strong means in natural language generation and in Machine Translation to rank sentences and to grade their well-formedness, fixation n-gram models could be instrumental in classifying reading behaviour and to decide, for instance, in what consists fluent reading and when reading difficulties occur.

7   Conclusions

The paper describes ongoing work to integrate the output of two user monitoring programs which produce fundamentally different kinds of User Activity Data (UAD): while TRANSLOG produces a fine-grained resolution of keyboard activities, GWM records eye-movements, computes fixations and gaze-to-word mapping hypotheses.

The integration of these data is a prerequisite for the design and implementation of a query language which aims at linking product data with process data, to retrieve and compare patterns of UAD across different texts and users. The paper investigates objectives and potentials for such a query language and gives a few examples of retrieved patterns.

References

1. Carl, Michael, Arnt Lykke Jakobsen, Kristian T. H. Jensen: Studying Human Translation Behavior with User-activity Data. NLPCS 2008: 114-123
2. Engle, R. W., Tuholski, S. W., Laughlin, J. E., Conway, A. R. A (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. Journal of Experimental Psychology: General, 128, 309-331.
3. Heitz, R. P., Schrock, J. C., Payne, T. W., Engle, R. W. (2008) Effects of incentive on working memory capacity: Behavioral and pupillometric data. Psychophysiology, 45, 119-129.
4. Kahneman, D. and Beatty, J. (1966). Pupil diameter and load on memory. Science, 154, 1583-1585.
5. Krings, H. P. (1986) Was in den Kpfen von bersetzern vorgeht. Narr: Tbingen.