20. PERCEIVING AND PRODUCING THE FROG STORY

KENNETH HOLMQVIST*, JANA HOLSANOVA*, VICTORIA JOHANSSON**
AND SVEN STRÖMqvIST

*Department of cognitive science, Lund University
**Department of linguistics, Lund University

INTRODUCTION

Consider the screen-shot in Figure 1. It is derived from a recording of a computer-based narrative task, using the word-less picture story *Frog where are you?* (Mayer, 1969) as elicitation instrument (see also Berman and Slobin 1994). The screen-shot was taken 2 minutes and 58 seconds into the recording. At that point, the subject, a 23-year old Swedish university student, had just finished writing in relation to the first picture (out of twenty-four) of the story.

How can we get a handle on the processes behind the product in figure 1? How did the flow of writing interact with the distribution of visual attention? These are the main questions addressed in the present paper. Starting from some preliminaries in reading and writing research, we take the step to presenting a methodology and an analysis example from a new research project1 where computer logging of writing activity is combined with eyetracking to derive a profile of the interaction between picture viewing and writing during the production of a picture-elicited narrative.

SCRIPTLOG: A WINDOW ON WRITING THE FROG STORY

The data underlying the present analysis was collected by means of the computer tool *ScriptLog*. *ScriptLog* is developed specifically for the purpose of studying the process of online writing (Strömqvist and Malmsten, 1998; Strömqvist and Karlsson, 2002;

1 "The Dynamics of Perception and Production during Text Writing" (project leaders: Kenneth Holmqvist and Sven Strömqvist; sponsored by Vetenskapsrådet, Sweden’s National Research Council).
Figure 1. Screen-shot from a writing experiment with a 23-year old Swedish university student. Stimulus picture (upper left), the resulting narrative text fragment in the editor window (right), and the recording panel (bottom left).

see also www.ScriptLog.net). ScriptLog keeps a record of all events on the keyboard, the screen position of these events and their temporal distribution. From a ScriptLog record, you can derive not only the finally edited text from a writing session, but also the linear text with its temporal patterning, pauses and editing operations. A writing session can also be played back, as is illustrated in the screen-shot in figure 1.

Further, ScriptLog allows you to design writing experiments with elicitation stimuli. In the example reported here, the picture booklet "Frog, where are you?" (Mayer, 1969) was administered as such an elicitation stimulus. During the writing task, the subject activates each new picture from the frog story by clicking a button labelled "Next". The button is in the same location as the floating recording panel in figure 1.

VISUAL ATTENTION FROM PICTURE TO NARRATIVE

Writing a narrative on the basis of a picture stimulus is a complex task. Visual information from the pictures has to be taken in and interpreted for the construction of the narrative (Strömqvist and Dey, 1993: 137-141). For example, still pictures must be interpreted as events, and pictorial elements must be translated into linguistic code. The linguistic output must be organised on the level of lexical, grammatical and narrative
structure, and then written down (in our case on a keyboard). The written output (on the screen) must be monitored (and it is indeed often revised). In all of these processes, visual attention plays an important role.

When we look at a picture or text, we move our eyes across the text, but not smoothly as we may intuitively think. Instead, the eye makes a series of short stops interleaved by quick jumps. The stops are called fixations, and during a fixation we can see what we look at. The jumps are called saccades, and during a saccade, we are blind (cf. the reading pattern in figure 3).

We may think that we are at leisure to look at pictures any way we want to. In fact, there are often elements in a picture that attract more attention than others. Faces, and in particular eyes, virtually always attract more visual attention than other parts of a picture, since faces are so important to us.

"The observer's attention is frequently drawn to elements which do not give important information but which, in his opinion, may do so. Often an observer will focus his attention on elements that are unusual in the particular circumstances, unfamiliar, incomprehensible, and so on." (Yarbus 1967:191).

Usually we spend several successive fixations examining an element. These fixations cluster into what is sometimes called an area of interest. When we have finished examining an element, we sometimes look for other elements, but often return to the first one.

"(...) when changing in points of fixation, the observer's eye repeatedly returns to the same elements of the picture. Additional time spent on perception is not used to examine the secondary elements, but to reexamine the most important elements." (Yarbus 1967:193).

Our viewing behaviour is to a large extent determined by the particular task we are faced with. For instance, if we are asked to judge the relation between people in a picture, we will typically look back and forth between the people and their faces (Yarbus 1967). In a task where we are asked to create a narrative from a stimulus picture, the natural thing to expect is that elements which appear as agents in the narrative or which present cues to the narrative flow will attract most attention.

Holsanova (2001) asked her subjects to describe a picture verbally which they were simultaneously watching (see figure 2). Her data indicate that subjects start by looking at picture-inherent objects, units and gestals. As their picture viewing progresses, they tend to create mental units that are more independent of the concrete picture elements. They may make large saccades across the whole picture, picking up information from different locations to support concepts which are distributed across the picture (like "spring" or "friend"). With the increasing cognitive involvement, observers and describers tend to return to certain areas, change their perspective and reformulate or recategorise the scene. The dynamics of this categorisation process is reflected in the usage of many refixations in picture viewing and reformulations, paraphrases and modifications in picture description.
Holsanová found two different styles concerning how speakers described one and the same picture. Attending to spatial relations in the picture was dominant in the more static technical style, while attending to the flow of time was the dominant pattern in the more dynamic narrative style. In the technical style, the picture in figure 2 was decomposed into fields that were then described systematically, using a variety of terms for spatial relations. In the course of description, informants established an elaborate set of referential frames that were used for localisations. They gave a precise number of picture elements, stated their colour, geometric form and position. Typical of the technical description style was a frequent use of nouns, existential constructions ('there is', 'it is', 'it was'), auxiliary verbs and passive voice.

In the narrative style, subjects focused primarily on the dynamic events in the picture. They followed a narrative schema: their description started with an introduction of the main characters, their involvement in various activities, and a description of the scene. This introductory part resembled the first phase of a narrative (Holsanová 1986, Labov & Waletzky 1973, Strömquist & Day 1993). Although there was no temporal or causal order inherent in the picture, speakers explicitly marked that they were talking about steps in a process, about successive phases, about a certain order. The dynamic quality of this style was achieved by a frequent use of temporal verbs, temporal adverbs and motion verbs in active voice. Discourse markers were often used to focus and refocus the picture elements, and to interconnect them.
THE FLOW OF WRITING A NARRATIVE

The flow of writing, as evidenced in a ScriptLog recording, reflects both lower level processes (such as spelling and lexical retrieval) and higher level processes (such as the construction of a background, a story line, a coda) in text production (Wengelin 2002). With respect to higher level processes, Johansson has shown personal narratives typically to be characterized by long pauses clustering at the beginning and end of the writing session with more extended phases of fluent writing inbetween (Holmqvist et al. 2001; Johansson in preparation). The fluent middle phases reflect the relative ease of retrieving and formulating the narrative events which make up the central part of the story line, whereas the relatively hesitant initial and final phases reflect the more complicated decisions involved in constructing a background to the story and some kind of punch or moral.

Strömqvist et al. (in press) used ScriptLog to study writing development (three age groups: 9-12-year-olds, 15-year-olds and adults) as reflected in keystrokes, pauses and editings across the discourse (24 pictures) of Mayer’s Frog, where are you? (Mayer, 1969). They found, that, on a group level, the 9-12-year-olds wrote the least, the 15-year-olds a little more, and the adults the most. Further, the profiles indicated group differences in terms of the distribution of writing activity across the discourse. The 9-12-year-olds wrote a lot in the beginning, then less and less and they had a very meagre finish. It is reasonable to interpret this profile as indicative of effort and exhaustion. In contrast, the 15-year-olds showed a much more smooth and balanced profile. Along the same line of interpretation, they seemed to be better able to plan their discourse and to control their expenditure of effort. They also spent a little more on the finish, than the younger children. The adults, finally, described a very dynamic overall profile: they spent a lot of their writing activity in the beginning and at the end of the story and varied themselves inbetween according to the richness and ramifications of the individual pictures.

Further, Strömqvist et al. found that the 9–12-year-olds had the largest and the adults the smallest proportion of pauses in relation to amount of writing (as measured by number of keystrokes). Many of the 9–12-year-olds made a relatively short planning pause before they started writing in relation to the first picture of the frog story (on average, less than half a minute), the 15-year-olds tended to make longer discourse initial planning pauses (around 1 min), whereas many of the adults made very long planning pauses (on average, more than 2 min).

In terms of editing activity, the amount of editing tended to decrease towards the last third of the story for all three age groups. The last picture/episode constituted an exception in the two older age groups: the 15-year-olds and adults performed a relatively large amount of editing in relation to the last picture. The 9-12-year-olds, in contrast, performed very little editing in relation to the last picture. In terms of relative amounts, the 9–12-year-olds edited 23.3% of the keystrokes they spent on the initial picture, but only 10.1% on the final picture. The great majority of the editings concerned spelling mistakes, and there still remained many spelling errors in the final edited text. The 15-year-olds had a higher and more balanced proportion of editings:
30.6% for the initial and 26.6% for the final picture. Again, the great majority of the editing concerned spelling mistakes, but there remained few errors in the final edited text. The adults tended to have a balanced editing rate, on average 15.5% for the initial and 15.6% for the final picture. The great majority of their editor’s concerns about spelling, and their texts were almost completely error free from the point of view of spelling. Thus, both the amount and nature of the editing operations were subject to developmental change.

In another study (see Aisenman 1999a, 1999b), ScriptLog recordings of text writing in different genres by school children and university students were analysed. Among other things, the temporal patterns and editing patterns were related to the progression of information in discourse (Strömquist, 2000). It was found that more extensive editing operations were only initiated in major information boundaries (sentences, paragraphs) in the text. In contrast, longer pauses were distributed both in larger information boundaries and between words within syntactic phrases.

What are the writers doing during the pauses? Are they reading their own text? Are they looking at the pictures? Are they monitoring self corrections or revisions? Information about the writer’s eye movements would help us refute some of these hypotheses and tease out what the writer is doing.

THE READING PROCESS

During the production of a written narrative, there is often a reason to read the text produced. When we read, there is a conventionalised order in which we look. Figure 3 shows a young woman reading a short passage of Pippi Longstocking in Swedish. She has already read six lines of text, but when she reads the words “fast jag kom inte ihåg” (“although I do not remember”), the reader herself obviously does not remember who the referent of “I” is. She make two saccades upwards and one to the left, finding “Tommy”, which is the correct referent. She then goes on reading.

---

Figure 3. Fixations (circles) and saccades (lines) in a text. From Noëland (2000).
Fixations typically measure from 100 to around 400 ms. Saccades take 20–50 ms to complete. That is, the average adult reader makes 3–5 fixations each second. It is only possible to read text which is at, or very close to, the fixation point, within approximately 2° of visual angle. Due to the uneven distribution of the different photoreceptors on the retina, text further out in the visual field cannot be read without making a saccade to it.

One line of classical reading studies focuses on issues such as which words are skipped (typically small, frequent words like ‘the’ and ‘and’), and where inside a word the eye typically lands (typically some 30–40% into the word). Several studies have looked at our ability to use information from words which we have not yet fixated, the so-called preview benefit effect. This line of research is today summarised in the E-Z reader model; for a good overview see Reichle, Rayner and Pollatsek (2000). Other eye-tracking studies have looked at dyslectic readers (e.g. Rayner 1978), at speed-reading (Yarbus 1967) and individual reading strategies (Nordlund 2000, Hyönnä et al. 2002). However, the reading process during the production of narratives has not previously been investigated.

**PERCEIVING AND PRODUCING THE FROG STORY: AN ANALYSIS EXAMPLE**

As mentioned above, one of the output files from a ScriptLog analysis is the *Linear file*, which renders all keyboard and mouse events during the writing session in exactly the order they happened. From a linear file the researcher gets an overview over the progression of the writing production at hand, without having to play-back the ScriptLog recording. An example of a linear file is shown in figure 4. Information about movements, such as using the backspace and delete keys, starting and ending

---

<START><SECTION1><STIMULUS-ONSET><0.12.519>Det är precis innan läggsdags och <0.06.980>Rutger och hans <0.06.810>trogne vän <0.14.495>Buster<0.07.341> <0.09.760>inspekterar dagens fångst. De har tillsammans fångat en grön liten groda i dammen.<BACKSPACE>sjön. För säkerhets<BACKSPACE> skull har de lagt honom i en glasburk så att han inte skall kunna rymma.<0.16.757> De är båda mycket nöjda med dagens arbeite.

<START><SECTION1><STIMULUS-ONSET><0.12.519>It is just before bedtime and <0.06.980>Rutger and his <0.06.810>faithful friend <0.14.495>Buster<0.07.341> <0.09.760>are inspecting today's catch. They have together caught a green little frog in the pond.<BACKSPACE>lake. For the sake<BACKSPACE> of security they have put him in a glass jar so that he won't be able to run away.<0.16.757> They are both very pleased with today's work.

---

*Figure 4. A linear file (ScriptLog) showing writing behaviour in relation to the first picture of the frog story.*
the writing session, activating various stimuli, and pausing (shown in minutes, seconds and milliseconds) is indicated with angular brackets. The linear file differs from the final edited text (shown in figure 1), in that the former contains elements that have been deleted from the final edited text. In the linear file, the insertions are thus found in the order in which they were produced during the writing session. The Swedish linear text is followed by a facsimile in English.

The first item in the linear file shown in figure 4 is "<START>" which indicates that the subject pressed the start button to activate the editor window and the first stimulus picture. "<SECTION1>", then, indicates that the first elicitation picture was loaded and "<STIMULUS-ONSET>" that it was displayed on the screen. Pauses are indicated within angular brackets.

From a ScriptLog recording we can further obtain information about deletions, revisions and pausing time during a writing session. In the linear file in figure 4, only pauses longer than 5 seconds are shown. Deletions are indicated by "<BACKSPACE>". In the first instance, "<BACKSPACE>", the writer performs a lexical replacement. She strikes backspace seven times to delete the word dammen (‘pond’) and a period, and then she writes the word sjön (‘lake’) and a period. In the second instance, "<BACKSPACE>", the writer corrects a spelling/writing mistake by deleting an s in säkerhets (‘security’s’). She replaces the s with a space and then writes skall (‘take’). In the final edited text, one spelling mistake remains uncorrected: the word inspektorer (should be inspektör ‘inspect’).

Further, we see that the writer had seven pauses longer than 5 seconds. The first pause of 12.5 seconds occurs in the beginning of the text. One can hypothesise that she looks at the picture, planning what to write next. Two of the pauses precede the naming of two story characters (the boy Rutger and the dog Buster), probably reflecting the writer’s effort trying to find suitable names for these characters. One pause precedes the noun phrase trogne vän (Buster) (‘faithful friend (Buster)’), and two consecutive pauses precede the verb phrase inspektorer dagens fängst (‘are inspecting today’s catch’). In both of these cases we assume that the pauses occur for planning reasons, such as composing the adjective phrases, selecting the appropriate verb (inspektöra) etc. The longest pause (16.7 seconds) occurs in the sentence boundary before the last sentence. During that pause, the subject is probably reading through what she has written so far, in order to see if something is missing, or to get new ideas on how to continue the story. Before moving on to the next picture, she writes a final sentence, summing up and evaluating: De är båda mycket nöjda med dagens arbete (‘They are both very pleased with today’s work’).

Our detailed analysis example of visual behaviour and writing behaviour focuses on the first 90 seconds of the frog story narration. On the basis of eye-tracking and ScriptLog data, we have created a temporally ordered multimodal score-sheet (defined in Holsanova, 2001) which shows the visual behaviour and writing behaviour synchronised over time. Figure 5 shows the score-sheet for those first 90 seconds.

In figure 5, the time line is projected downwards from the top to the bottom of the page. Objects present on the screen in the original writing condition (see figure 1) are represented as columns in the score-sheet. The rightmost column represents the
Figure 5. A multimodal score sheet representation of visual behaviour and writing behaviour during the first 90 seconds of a frog story narration.
editor window and the second column from the right represents the picture (depicting the boy and his dog watching the frog). Then, in the order from right to left, follow columns representing details (sub-areas) of the picture: the boy, the frog, the dog, the window in the room, the lamp and bed. The left-most column in figure 5, finally, represents the "Next" button (floating panel). When the subject is looking at a given object/sub-area of the screen for a certain period of time, this is indicated through a corresponding shading/blackening along the time line in the column representing the object/sub-area in question. For example, during the first few seconds, the subject is first looking at the lamp and bed in the stimulus picture, then at the frog and the boy, and then at the "Next" button. With the exception of the "Next" button, the viewing of these different objects are conflated in the Picture column. A grey stripe across all columns indicates that the subject is looking in another direction than that of the screen (for example, at the keyboard, when typing).

To the right of the picture, a layer with the writer's emerging text is added. Each letter has been placed at a height that indicates the point in time when it was typed. In effect, the decline of the row of letters mirrors the speed of writing: the decline is small when typing is fast and increases as typing gets slower. Consider, as an extreme case in point, the steep decline between peci and s as the subject writes the word peci ("just/precisely"), making a 2 second pause before the s at the end of the word. Pauses in writing which are longer than 5 seconds are marked with angular brackets.

From the temporally aligned eye-tracking and writing data in figure 5, several patterns can be deduced and interpretations suggested. Thus, during the 12.5 seconds pause preceding the first keystroke of the writing activity, the eye-tracking data shows that the writer is looking at the picture. For the greater part of this time, she is looking back and forth between the three agents: the dog, the frog and the boy. This can be interpreted as time spent forming the first ideas for the narrative.

Further, typing activity in the subject always coincides with grey stripes across the columns in figure 5, indicating that the subject is consistently looking at the keyboard when writing. When she redirects her visual attention to the screen, she very often looks directly at the text in the editor window (see rightmost column in figure 5). We interpret these latter fixations as indicative of a need for visual feedback in the writing process.

The naming of the story characters is associated with very long pauses. During the almost 7 seconds long pause around 20 seconds into the writing activity, the writer first looks at the editor window, reading the text she has produced so far. She then alternates between looking at the boy and the dog, before she looks back at the editor window and types a name referring to the boy, Rutgers, together with the beginning of the noun phrase ach hans ("and his"). Then there follows a 6.8 seconds long pause, during which the subject divides her visual attention between first the dog, the boy and the frog, then the editor window, the keyboard, the editor window again, and ends with a long look at the dog. After the pause, she types the noun phrase tøgne røn ("faithful friend"). In relation to both of these pauses the visual data indicates that she is searching the picture for information which might help her formulate the name and the descriptions.
The naming of the dog seems to be an even more effortful procedure. The writer spends the better part of an almost 14.5 seconds long typing pause looking at the dog. It is possible that the image of the dog helped her finding a suitable name, but it is more likely that the persistent looking at the dog helped her focus on the task of finding a name. After the pause, the subject types the name *Buster*. After *Buster*, the subject makes two consecutive pauses (7 and 9 seconds, respectively). She spends them looking back and forth between the dog, the frog and the boy, except for an almost 5 second long look down at the keyboard. Maybe she is looking down in order to start typing, but then changes her mind and decides to take a second look at the picture first. Again, this second look, around 75 seconds into the writing session, is directed towards the three main characters of the narrative. The subject then types *inspektor dagens fängst* ('inspecting today's catch'). Two very short glimpses at the text in the editor window are interleaved with this typing activity.

**RESEARCH DIRECTIONS**

The behavioural flow of writing revealed by ScriptLog methodology takes the analyst closer to the processes brought to play when a text is being produced. When eyetracking data is added, an enhanced picture of attentional processes during writing emerges. The combination of computer logged writing with eyetracking offers a window on the dynamic interplay between perception and production during text-writing.

Many questions and research problems remain to be solved, before a more fullfledged picture of textwriting emerges from our budding paradigm. In our further research, the subactivity of text revision will receive a special focus, since this is an activity which necessitates an interaction between reading and writing. It therefore provides a particularly fortuitous window on the dynamic interplay between perception and production during textwriting. Further, extensive empirical research needs to be carried out, in order to determine developmental patterns and patterns characteristic of writers with reading and writing problems. We believe that, once these patterns are mapped out, the analysis of textwriting online has an important role to play for the enhancement of tests and diagnostic procedures, of language pedagogy and computer based writing support tailored to users with different needs and different abilities (c.f. Ahlén and Strömqvist 1999).

For example, in the European project COST A8 "Cross-linguistic studies of discourse level writing in dyslexics" (see, e.g., Olofsson & Strömqvist 1998), we analysed ScriptLog recordings of adult dyslexics and groups of adult controls across English, Finnish and Swedish. It was found, among other things, that the dyslexics had a lower production rate and made more spelling and interpunctuation errors than the controls (Erskine 1999; Wengelin and Strömqvist 2000; Wengelin 2002). Wengelin and Strömqvist (2000) also found the dyslexics to have a larger proportion of both editings and pauses within words as compared to the controls. Further, Wengelin and Strömqvist found that the dyslexics tended to make longer pauses before terminating a sentence with a major delimiter as compared to when they were about to start a new sentence. The reverse was true of the controls. Wengelin and Strömqvist suggest
that the longer sentence final pauses in the dyslexics are indicative of a more effortful monitoring of the sentence (reading and checking the sentence before turning to the production of the subsequent sentence). An experiment which combined ScriptLog with eyetracking would make it possible to test this hypothesis.

In another project, directed by Ruth Berman, "Developing literacy in different contexts and in different languages" (see Aisenman 1999a; 1999b), ScriptLog recordings of text writing in different genres by school children and university students were analysed. In a case study, the temporal patterns and editing patterns were related to the progression of information in discourse (Strömquist, 2000). It was found that more extensive editing operations were only initiated in major information boundaries (sentences, paragraphs) in the text. In contrast, longer pauses were distributed in larger information boundaries as well as between words within syntactic phrases. Here, information about the writer's eye movements would help us tease out what the writer is doing in the pauses in these different linguistic contexts. Are the long pauses in major information boundaries associated with a cognitive focus shift, for example, from writing a sentence to (re)reading an earlier part of the text so far produced, whereas pauses in an initiated but not yet completed sentence tend to be associated with visual information search of a kind which does not compete with the focus on the not yet completed linguistic structure? Our analysis of the passage in Figure 5 provides evidence in this direction.

Many pedagogical applications presuppose an awareness on the part of the student/writer. To what extent are writers consciously aware of the decisions they are making online as they are constructing and reconstructing their texts? An offline method for tapping such information from writers is the so called Think Aloud Protocol (see, for example, Bereiter and Scardamalia, 1987; Hayes and Flower, 1980). A disadvantage with this method is that the protocol intervention inevitably disturbs the production process (Janssen, van Wae & van den Bergh, 1996). A post-writing debriefing interview supported by playback of a ScriptLog recording may prove to be a way around that problem.

Another line of research relates to our growing archive of ScriptLog recordings of writing activity. To date, we have some 2000 recordings from 10 different languages and from writers of different abilities. The archive has already been used as a resource for the construction of a probabilistic tool for spelling support for dyslexic writers. In the near future, the archive will also serve as a testing ground for new types of crosslinguistic research questions. For example, do certain sequences of letters or certain grammatical constructions tend to be written faster or monitored more carefully in certain linguistic communities than in others? Again, we believe that the answer to this kind of questions can have important implications for applied areas, such as translation or second language learning.

REFERENCES


II. A developmental perspective on language and discourse