

# Cognition, multimodal interaction and new media<sup>1</sup>

Jana Holsanova<sup>2</sup>

Cognitive Science Dept., Lund University

jana.holsanova@lucs.lu.se

ABSTRACT: From a cognitive perspective, this paper summarises a number of theoretical and applied studies conducted by my colleagues and myself on the topic 'interaction with new media'. The focus lies on the users' behaviour: visual information gathering, interaction with the multimodal interface, browsing strategies and attentional processes during hypertext navigation. In addition, we also look at users' expectations and attitudes towards the medium. There are several methods that can be used in order to describe user behaviour and postulate a number of underlying cognitive mechanisms. In the following, I will show how eye-tracking data supplemented by simultaneous or retrospective verbal protocols, keystroke logging, and interviews can help us to investigate users' behaviour, the

---

<sup>1</sup> The paper is based on my presentation "Tracking multimodal interaction with new media" at the workshop on *The citizens use and comprehension of information on the Internet* arranged by Kerstin Severinson Eklundh and Britt-Louise Gunnarsson, Uppsala 18-19 June 2004.

<sup>2</sup> Vsechno nejlepší, Wlodku, k narozeninám a díky za vsechny společné diskuze a rozhovory. - Dziękuję bardzo. Z pozdrowieniami, Jana.

rationality behind this behaviour, and users' attitudes and expectations. The clusters and integration patterns discovered in empirical studies can be used in developing a new generation of multimodal interactive systems within human-computer interaction.

### **User Behaviour on the Internet**

The first study conducted in 1996 by David de Léon and myself was a qualitative study of user behaviour on the World Wide Web concerning hypertext navigation and browsing strategies. Eight participants were filmed while performing user-defined tasks and then asked to review the video-taped session during prompted recall. This data formed the basis for a series of descriptions of user behaviour and the postulation of a number of underlying cognitive mechanisms. Our results indicate that users lack ready made search strategies, prefer alternatives that are visible, immediately available and familiar, choose the path of least resistance, exhibit social forms of behaviour, engage in parallel activities, object to misleadingly presented information, have trouble orienting, are late in using appropriate strategies, are sensitive to matters of time, and are emotionally involved in the activity. Finally, we discuss how these results can contribute to our understanding of hypermedia (for details see de Léon & Holsanova 1997; this paper - as well as most of the following papers mentioned - can be downloaded in PDF format at <http://lucs.lu.se/jana.holsanova>).

### **Eye Movements and Attention**

In the following studies, we wanted to follow visual scanning and attentional processes more exactly and thus started using an eye tracker.

The motivation of eye movements is to bring to a particular portion of the visible field into high resolution so that we can see it in fine detail. When we focus our concentration and eye movements on a point, we also divert our attention to that point. Data on visual behaviour can thus be used as measures of cognitive processes. Eye tracking allows us to register precisely what is fixated and when. We can follow along the path of attention deployed by the observers and get insights into what the observers found interesting, what drew their attention, how a scene was perceived etc.

### **Picture Viewing and Picture Description**

In my dissertation (Holsanova 2001), I investigated visual scanning and simultaneous verbal description of complex scenes. The goal was to compare the contents of verbal and visual 'attentional spotlight' and to answer the following questions: Can we identify comparable units in visual perception and in language production? Does the order of units in the verbal description reflect the general order in which information was acquired visually? Is the content of the units in picture viewing and picture description similar? In order to synchronise the verbal and the visual streams, I created an analytic format called multimodal time-coded score sheets. With the help of this new analytic format, I could analyse temporal and semantic relations between speech and gaze and extract configurations of verbal and visual clusters from the synchronized data. (Concerning the results of my studies, see Holsanova 2006 or consult <http://lucs.lu.se/jana.holsanova>). The method combining eye tracking and verbal protocols in order to get an enhanced picture of attentional processes

has been further developed in our recent study (see The dynamics of perception and production during text-writing).

### **Picture Viewing and Mental Imagery**

It has been proposed that we use mental imagery when we mentally invent or recreate personal experience, when we read novels, plan future events, retrieve information about the physical properties and relations of objects, imagine transformations by mental rotation and mental animation and when we solve problems. Finke (1989:2) defines mental imagery as “the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event, either in conjunction with, or in the absence of, direct sensory stimulation.” A large number of eye tracking studies have been conducted to test mental imagery. In order to extend the finding of previous studies, we chose pictures of high complexity as our visual stimulus and conducted a number of new eye tracking studies (Johansson, Holsanova and Holmqvist 2005, 2006). Our data indicate that even for a complex picture, spatial locations are to a high degree preserved when describing it from memory. Our results suggest that subjects visualize the spatial configuration of the scene as a support for their descriptions from memory. The effect we measured is strong. The results can be interpreted as further evidence that eye movements play a functional role in visual mental imagery and that eye movements indeed are stored as spatial indexes that are used to arrange the different parts correctly when a mental image is generated. However, there are alternative interpretations: According to an ‘embodied view’, subjects leave behind ‘deictic pointers’ to locations of the scene in the environment,

which later may be perceptually accessed when they are needed. According to the 'perceptual activity theory' account, imagery is the re-enactment of the specific exploratory perceptual behaviour that would be appropriate for exploring the imagined object as if it were actually present.

### **Newspaper and Net Paper Reading**

Eye tracking has also been used in applied studies of readers' interaction with newspaper and net paper layout. For designers, the most interesting issues are entry points (where do the readers start reading?), reading paths (how do the readers navigate through the medium?), reading depth (how carefully do they read the articles?) and local design factors (the effect of colour, pictures, headlines, drop quotes etc.). Two behaviours can be distinguished: the reading behaviour (a well defined movements across the text) and the scanning behaviours (large saccades in almost any direction when the reader is evaluating articles: are they worthy of deeper processing?). We conducted several applied studies on newspaper and net paper reading (see references) but apart from the issues mentioned above, we also wanted to understand some of the underlying rationale and motivation of the behaviour. We were interested in readers' reflections, experiences, comments and attitudes towards the new medium. To achieve this aim, we used a combination of three methods: (i) eye tracking, (ii) retrospective verbal protocols supported by the replay of the interaction and (iii) interview data.

Let us review some of the main questions and answers. How do readers interact with net papers compared to newspapers? Does the new medium

influence our way of searching for news? How do readers orient, navigate, which reading paths and entry point do they choose? What are their attitudes towards the medium? Our results show that in fact net paper readers scan more and read less than newspaper readers. We furthermore investigate whether this result can be explained by the difference in layout, navigation structure and purpose of reading between the two media.

The reading and interaction patterns in the two media differed. Net paper readers seemed to have problems with orientation in the hyper-structure. Below are some of the comments that the participants made on their interaction which could be used as explanations of the patterns found on the basis of eye tracking.

Reading situation, purpose of reading

"I just have a quick look at the net paper during a break to see what has happened."

"It's a nice relaxation to spend time reading a newspaper over a cup of coffee."

Navigation

"I start from the first page to get an overview, but I have problems getting back."

Entry points

"I think it's difficult to decide whether to enter an article on the basis of the headline only. It easily happens that you misunderstand it."

"The content links were messy and it took some time to scan them through in order to find what to read."

Topic spectrum

”I think net papers are good because it is easy to compare different news sources and how they report about the same news.”

The reading situation and purpose of reading vary a lot between the different media. The comments on navigation confirmed the patterns in reading paths - net paper readers have difficulties orienting and navigating. In addition, switching to another story is very slow in net paper reading compared to newspaper reading. While it takes only 20-50 ms to move the eyes to another article on the newspaper fold, for net paper readers it takes much longer. They have to determine which article they want to read by reading content summary (link), they have to click on the link and wait for the article page to be downloaded. The feed back is very slow. Since the investment of switching story is such a high threshold, the readers rather avoid switching between stories. Once they decided to read a story, they read it to the end. To sum up, newspaper reading is characterised by a broad topic spectrum and shallow reading. Net paper reading, on the other hand, is characterised by a narrow topic spectrum and deep reading. For more results consult e.g. Holmqvist et al. (2003) and Holsanova & Holmqvist (2004).

### **Discussing Theoretical Frameworks**

Researchers within the sociosemiotic framework have suggested how a newspaper layout could be analysed (Kress & van Leeuwen 1996). This analytical framework has also been applied for the Internet (cf. Karlsson & Ledin, 2000). In one of our studies, we decided to extract hypotheses about reading of newspaper layout from Kress & van Leeuwen’s work and discussed them in connection to data from readers' authentic interaction

with a newspaper. In particular, we used eye tracking to empirically test hypotheses about entry points and reading paths. First, newspaper layout was analysed according to the sociosemiotic approach (without any knowledge of the actual reading behaviour). Second, eye movement data on the newspaper fold was analysed in three different ways: i) the temporal order of the attended areas was calculated in order to determine reading priorities; ii) the amount of time spent on different areas was calculated in order to determine which areas have been read most; iii) finally, reading depth was calculated in order to determine how carefully those areas have been read. The results show that the reading behaviour is very dynamic. Readers register the units predicted by the sociosemiotic theory but the reading paths are created in very different ways (cf. Holsanova, Rahm & Holmqvist 2003, 2006).

### **Dynamics of Perception and Production in On-line Text Writing**

The process of visual information gathering and on-line text production was studied in a current project in co-operation between researchers from the Cognitive Science Dept. and Linguistics Dept. at Lund University (project leaders: Sven Strömquist and Kenneth Holmqvist). Our project team collected data from 96 participants (balanced for age, gender and dyslectics/controls) and used an extended methodology: production-rate data from key-stroke logging (ScriptLog), eye-tracking data (iView) and follow up debriefing interviews supported by the playback of the interaction. For analysing the interaction between writing and gaze behaviour, we have developed an analytical tool (Andersson et al., 2006), inspired by the multimodal time-coded score sheets developed by



Holsanova (2001). The tool helps with analysing temporal and semantic synchrony in picture viewing and picture description. This offers an enhanced picture of the attentional processes: which objects or areas were scanned visually and which objects or areas were described verbally at a certain point of time. When investigating picture descriptions, we can see how the writers' attention is distributed between the stimulus picture, keyboard, computer monitor and elsewhere during writing. In sum, this analytic format gives deeper insights into the dynamics of perception and production during on-line text writing.

### **Integration and Interpretation of Text and Pictures**

Multimodality has been studied within many different disciplines: semiotics, sociosemiotics, textlinguistics, interface design, and human-computer interaction (cf. Holsanova 1996, for an overview, cf. Holsanova 1999, 2002). However, very few studies concern the way users perceive the interplay between the modalities: From which source do users' acquire information and how do they integrate it? Some first steps have been carried out with reading of diagrams and within applied studies of advertisements (Rayner et al. 2001). The process of semantic interpretation during picture-elicited writing will also be an important contribution to this topic (Holsanova 2003). To sum up, perception and interpretation of multimodality is strongly understudied and deserves more extensive focus in future research.

## **Reading Information Graphics**

A couple of recent studies at the Humanist Laboratory in Lund are devoted to the reading of information graphics. The reading paths between headline, text and infographics can give us hints about the integration and interpretation processes. The choice of entry points can tell us more about the readers' interests and the process of meaning-making. In one of our studies, we address the issue of whether the overall design of information graphics influences reading behaviour (Holmqvist et al. 2006, Holsanova et al., 2006. forthc.). Our second study focuses on factors that have effects on the comprehension of information graphics (Holmberg et al. 2006). The studies are relevant for the research on cognitive processes involved in multimodal integration, for the applications in printed and digital media and for the design of user-friendly multimodal documents. Currently, we are investigating the integration of text, pictures and graphics within the project "Information Graphics and Eye Movements. Tracking reader's interaction with a multimodal document genre", financed by the Helge Ax:sson Johnsson Foundation (project leader: Jana Holsanova, in cooperation with Nils Holmberg and Kenneth Holmqvist).

## **Summary and Conclusions**

Gaze behaviour reveals the path through the text, paths through the picture and the connections created between picture and text. It can be used to track attentional processes. If eye-tracking data is supplemented by simultaneous or retrospective verbal protocols, the analysis can give a very detailed view on reader interaction with the media. By combining various methods, such as eye tracking, keystroke logging, retrospective verbal

protocol supported by the replay of the interaction, and interviews, we obtain multiple 'windows into thought and action'. We can track users' navigation and orientation, detect interaction problems, evaluate interface design, test theoretical models against empirical results, investigate the perception of multimodality. It allows us to investigate user behaviour, the rationality behind the behaviour, and users' expectations and attitudes.

### **Visions**

One of the visions is to investigate individual 'reading styles' – a question that is currently studied within the project "Reading paths and reading styles on a newspaper spread"<sup>3</sup> at the Humanist laboratory at Lund University (project leader: Kenneth Holmqvist, in cooperation with Nils Holmberg, Henrik Rahm and Jana Holsanova). Another vision is to answer questions concerning cultural differences in reading text, pictures and graphics. Cognitive aspects of multimedia have been focused by Mayer (2005) and others. The connection between cognitive and cultural aspects has, however, not yet been addressed. A third vision is to track natural integration of different modes of communication (speaking, pointing, scanning pictures, writing text, etc.). When we have two or more continuous streams of behaviour, what does the semantic unification process look like? Where and when do we integrate those streams? Are

---

<sup>3</sup> "Läsvägar och lässätt på tidningsuppslaget – ett tvärvetenskapligt projekt med semiotiska analyser och ögonrörelsemätningar", financed by Erik-Philip Sörensens stiftelse.

they overlapping or sequential? Which are the meaningful units of our interaction? Oviatt (1999) stresses the need of guidance from cognitive science on co-ordinated patterns of human perception and production based on empirical evidence. The clusters and integration patterns discovered in such kind of empirical studies could be used in developing a new generation of multimodal interactive systems in human-computer interaction.

## References

- Andersson, B., Dahl, J., Holmqvist, K., Holsanova, J., Johansson, V., Karlsson, H., Strömquist, S., Tufvesson, S., and Wengelin, Å. (2006): Combining keystroke logging with eye tracking. In: Luuk Van Waes, Marielle Leiten & Chris Neuwirth (eds.) *Writing and Digital Media*, Elsevier BV (North Holland), 166-172.
- de Léon, D. & Holsanova, J. (1997). Revealing user behaviour on the world-wide web. *Lund University Cognitive Studies 60*. Lund university.
- Finke, R.A. (1989). *Principles of Mental Imagery*, Cambridge, MA: MIT Press.
- Holmqvist, K., Holmberg, N., Holsanova, J., Tärning, J. & Engwall, B. (2006): Reading Information Graphics - Eyetracking studies with Experimental Conditions (J. Errea, ed.) *Malofiej Yearbook of Infographics*, Society for News Design (SND-E). Navarra University, Pamplona, Spain, 54-61.
- Holmqvist, K., Holsanova, J., Barthelson, M. & Lundqvist, D. (2003): Reading or scanning? A study of newspaper and net paper reading. In Hyönä, J. R., and Deubel, H. (Eds.), *The mind's eye: cognitive and applied aspects of eye movement research* (657-670). Elsevier Science Ltd.
- Holsanova, J. (2004): Tracking everyday cognition. Invited lecture at KVIT 2004, Linköping, 6-7 May 2004.

- Holsanova, J. & Holmqvist, K. (2004): Med blick på nätnyheter. Ögonrörelsestudier av läsning i nätbaserade tidningar. (Looking at the net news. Eye tracking study of net paper reading) In: Holmberg, Claes-Göran & Svensson, Jan (red.), *Mediekulturer, Hybrider och Förvandlingar*. Carlsson förlag, 216-248.
- Holsanova, J. (1996). När säger en bild mer än tusen ord? Om samspelen mellan text och bild i gränssnitt. In: Löwgren, J. (red.) *Teman i människa-datorinteraktion*. 65 -76. Institution för Informatik, Lunds universitet.
- Holsanova, J. (1999). Olika perspektiv på språk, bild och deras samspel. Metodologiska reflexioner. (Different perspectives on multimodality. Methodological considerations) In: Inger Haskå & Carin Sandqvist (eds.): *Alla tiders språk*. Lundastudier i nordisk språkvetenskap A 55. (117-126). Lund University Press, Lund.
- Holsanova, J. (1999). På tal om bilder. Om fokusering av uppmärksamhet i och strukturering av talad beskrivande diskurs. (Speaking of pictures. Focusing attention and structuring spoken descriptive discourse in monological and dialogical settings). Licentiate dissertation. *Lund University Cognitive Studies* 78.
- Holsanova, J. (2001): Picture Viewing and Picture Description: Two Windows on the Mind. Doctoral dissertation. *Lund University Cognitive Studies* 83.
- Holsanova, J. (2006): Dynamics of picture viewing and picture description. In: Albertazzi L. (ed.), *Visual thought. The depictive space of the mind*. Part Three: Bridging perception and depiction of visual spaces. *Advances in Consciousness Research*. Benjamins, 233-254.
- Holsanova, J. (under utg.): Användares interaktion med multimodala dokument. I: Gunnarsson, B.-L. & Karlsson, A.-M. (utg.) *Ett vidgat textbegrepp*. TeFa rapport. Stockholms universitet.
- Holsanova, J., Rahm, H. & Holmqvist, K. (2003): Tidningarna och läsbarheten. I: *Grammatik i fokus*, volym I. Festskrift för Christer Platzack, 55-66. Institutionen för nordiska språk, Lunds universitet.
- Johansson, R., Holsanova, J. & Holmqvist, K. (2005): What Do Eye Movements Reveal About Mental Imagery? Evidence From Visual

- And Verbal Elicitations. In Bara, B. G., Barsalou, L., Bucciarelli, M. (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 1054-1059. Mahwah, NJ: Erlbaum.
- Johansson, R., Holsanova, J. & Holmqvist, K. (2006): Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science* 30:6, 1-27. Lawrence Erlbaum.
- Karlsson, Anna-Malin & Per Ledin (2000): Cyber, hyper och multi. Några reflektioner kring IT-ålderns textbegrepp. Human IT. URL: <http://www.hb.se/bhs/ith/23-00/amk.htm>
- Kress, G. & T. Van Leeuwen (1996): *Reading Images. The grammar of visual design*. London & New York: Routledge.
- Mayer, R. (2005): *Cognitive Theory of Multimedia Learning*. Cambridge University Press.
- Oviatt, S. L. (1999): Ten Myths of Multimodal Interaction. *Communications of the ACM*, Vol. 42, No. 11, Nov. 99, 74-81.
- Rayner, K., Rotello, C.M., Stewart, A. J., Keir, J. & Duffy, S.A. (2001): Integrating text and pictorial information: Eye movements when looking at print advertisements. Paper at the ECEM in Turku.