MAD 2010, Moissac, 17-20 mars 2010

Spatial Coding and information retrieval in multimodal documents

Claudio Vandi (1), Thierry Baccino (2)

(1) Laboratoire CHArt, Université de Paris 8 - LUTIN UMS CNRS 2809
vandi@lutin-userlab.fr
(2) LUTIN UMS-CNRS 2809
baccino@lutin-userlab.fr

Abstract In this paper we study the effects of Spatial Coding (Murray, Kennedy, 1987; Kennedy, 1992) on the way readers process semantic information while searching for a target in multimodal (text- image) documents. Our aim is to understand how spatial indexes are used by readers in information retrieval tasks, and what are the interactions between semantic distance and spatial distance. We present an eye tracking experiment in which 24 participants were asked to search for information in simple multimodal digital documents while we manipulated their ability to associate spatial indexes to semantic content and changed the location of the target text between the inspection and the search phases. The results show that participants' performances worsened with the increasing of the semantic distance between the target text and the text that took its place. Results also show that readers can rapidly adapt their information searching habits to the spatial organization of the document.

Keywords: spatial coding, multimodal, information searching, reading, eye tracking, habits.

1 Introduction

1.1 Spatial organization and semantic processing in texts

Spatial organisation can be a powerful tool for signalling text organisation and is likely to have a direct influence on information searching skills and text comprehension abilities of readers. Previous studies on reading and information retrieval put forward a hypothesis referred to as the "spatial coding hypothesis" (Kennedy, 1992; Kennedy et al., 2003), according to which "readers must maintain, and use, a level of representation of text that involves the computation of spatial coordinates" (Kennedy, Murray, 1987). Authors who support this hypothesis show, for example, that readers are capable of localising the first word of an anaphora by making a single and very precise saccade to its position, thus presupposing that they first *coded* this position, even if they weren't explicitly asked to do so (Kennedy, 1982; Murray, Kennedy, 1987). Another study shows the ability to code spatial position of words is a fundamental step in becoming an expert reader (Murray, Kennedy, 1988). Other studies show that readers that are able to associate semantic and spatial information have a better understanding of texts, compared to subjects that receive only non localised information (Virbel et al., 2005). This hypothesis is also consistent with "position special" theories of vision that claim that spatial features (the "where") play a special role in vision, compared to other visual features like colour and shape (the "what") (Van der Heijden, 1993; Van der Heijden et al., 1999; Spivey, 2001). Furthermore, the "spatial coding" phenomenon is also considered as one of the proof for the validity of the "extended mind" theory (Clark, Chalmers, 1998), according to which readers do not need to create a mental copy of the whole document while reading it, but just place some spatial indexes on the page in order to use it as an "external memory" that can be accessed to retrieve information when needed (O'Reagan, 1992; Richardson, Spivey, 2000).

1.2 Spatial coding and its relation to semantics in multimodal documents

Even if it is hard to find someone who denies that people retain *some kind* of spatial information while reading, to understand the role that spatial indexes can have on text signalling organisation and text comprehension we still need to answer a question concerning the *target* of this spatial coding. The main controversy in this respect concerns the role of semantic information, and can be expressed as follows: is spatial coding associated with the position of graphical elements (i.e. memory of the word's shape) or is it associated with the position of meaningful entities (i.e. memory of the semantic entity) ? For example in the following example from Kennedy and Murray (1987):

(1) The novels in the library had started to go mouldy with the damp. novels

When readers go back to the anaphora "novel" is their search aimed at the *meaning* of the word or do they just remember its *aspect* ? Our hypothesis (that we share with Baccino, Pynte, 1994) is that *spatial coding is semantically motivated* and consists in associating semantic labels with locations in space, so that actions (be they real actions or oculomotor ones) that are addressed to these locations will be semantically charged. This coding allows readers to develop dispositions towards action: having attributed a semantic value to specific locations in space, they know where to look for information and don't have to scan the entire scene before acting.

As far as the organisation of multimodal documents is concerned, our hypothesis about spatial coding can be articulated in three more detailed ones:

1. If spatial coding is based on spatial localisation of *semantic* information (i.e. not on its surface aspect) we should be able to observe its effects even if the cue and target do not share the same surface aspect (e.g. they belong to different modalities).

2. If spatial coding consists in attributing semantic labels to locations in space, we expect that if the semantic content of the text changes after the reader has labelled it, when he

will search for information again, he will behave in different ways if the new content is semantically associated or not associated to what he expects to find.

3. If readers rely on spatial coding for semantically organising a text, we expect that even when they are presented with documents, the spatial organisation of which is unconventional, they will adapt their research habits to the spatial structure of these documents.

Verifying these hypothesis would allow us to better understand how spatial and semantic relations influence the cognitive organisation of a text and how spatial and semantic information should be coupled if we want to create coherent and easy-to-search multimodal documents. This would be particularly important for the design of digital documents (web pages, digital documents for education) where document organisation changes in a dynamic way according to the support being used (desktop monitor, mobile device) or the contents being displayed (e.g. a news web page). To verify our hypothesis we performed an eye tracking experiment in which participants were asked to search for information in a series of simple multimodal digital documents composed of three texts and one image.

2 Methods

2.1 Participants and material

24 participants took part in our eye tracking experiment. They were divided into two equal groups: a *test group* and a *control group*. Both groups were presented with a series of 24 trials on a screen. For the test group, each trial was composed of two documents: the first one made up of three texts organised in a circle (left, top, right) and the second of an image that represented a common[1] object or animal in the middle of the screen surrounded by the three texts. For the control group, the first document presented only the image, the second document was composed of an image and three texts, identical to those of the test group. Each text was 320 characters long on average and described a common object or animal. In each document, only one of the three texts was strongly associated with the image. For the other two texts, one was only weakly associated with the image, and the other was non-associated. The intensity of this semantic association was computed using *Latent Semantic Analysis* (LSA) *one to many* comparison (Landauer et al., 1998; Bellissens et al., 2004).

2.2 Procedure

For the test group, in each trial, the documents were presented in two phases: an inspection phase and a search phase. During the inspection phase, participants had 10 seconds to skim the document. After ten seconds, a black screen appeared for 1 second and then the search phase began. During the search phase, the participants had to find the text that was strongly associated with the image. Participants were asked to locate this target text and press the spacebar. To make sure they located the target, they were asked to fix the correct text in their gaze while pressing the spacebar. Once they pressed it, they passed on to the next document.

1 Between 10 et 120 occurrences in the « Lexique » database http://www.lexique.org/listes

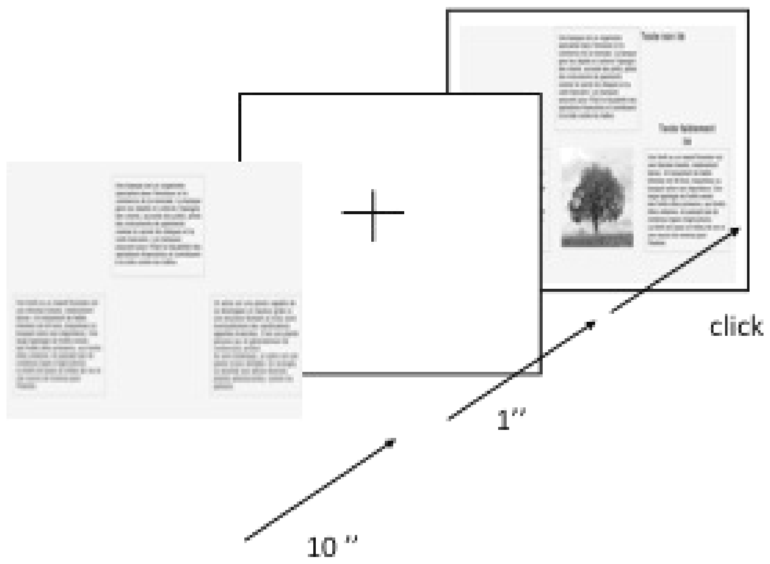

Figure 1 Schematic stimulus display for the test group. An inspection phase (10 seconds), followed by a neutral screen with a central cross (1 second) and a search phase (until the participant press the spacebar).

Between the inspection and the search phase, the content of the texts was the same. As far as their positions are concerned, in 50% of the cases it didn't change (a "No change" condition); in 25% of the cases, the target text swapped its position with the *weakly* associated test (a "Weak change" condition); in another 25% of the cases, it swapped its position with the *non-associated* text (a "Non-associated change" condition). The initial positions of texts were counterbalanced so that probability of the associated text appearing in any of the three positions was the same.

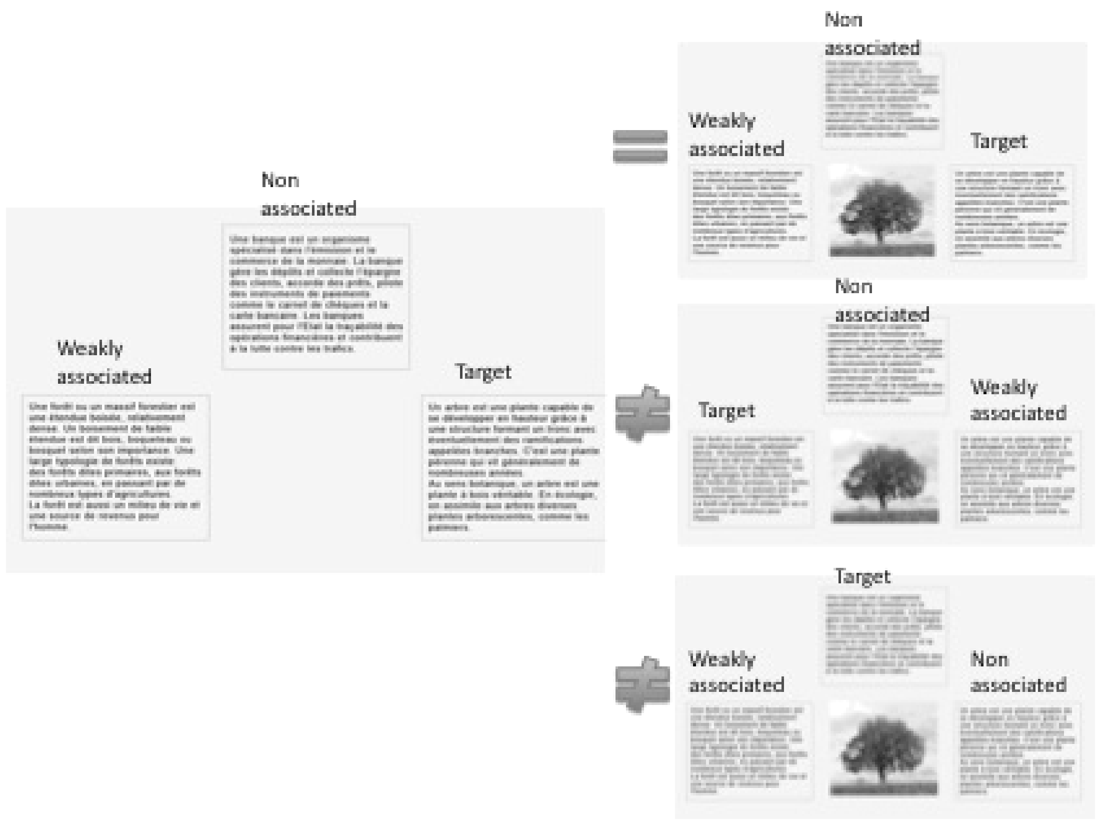

Figure 2 Schematic example of a trial for the test group. Left image : inspection phase. Right images : possible alternatives for the search phase. Target text: tree. Weakly associated text: forest. Non associated text: bank.

The procedure was the same for the control group, except that during the inspection phase, participants only saw the image, without texts. In this way, they didn't receive any priming about the spatial position of information during the inspection phase. During the search phase, the screens presented were identical to those of the test group. The task was also the same: during the inspection phase, the participants had to locate the associated text and press the spacebar. But since these participants had never seen the texts before the search phase, they could not experience any sort of spatial modification with respect to the inspection phase in which they saw just the image.

2.3 Eye-Movement Recording

Eye movements were recorded by means of a Tobii 1750 eye tracker. Response Times were recorded too. To record scanpaths, the screen was divided into four Areas Of Interest (AOIs) one for the central image, and three for each of the text zones (left, top, right).

3 Data Analysis

3.1 Error rate

For the scope of this study, an error was defined as each trial in which a subject pressed the spacebar while looking at the wrong text. The error rate (averaged over all participants) was 10% with no significant difference for the three conditions (no change, weak change and non related change) within the test group ($t(22)= -0.85$; $p =.96$*) and no significant differences between the test and the control group ($F(2,22)= 1.11$; $p = .34$*).

3.2 Eye-Movement Data

We analysed Eye-Movement and Response Time only for error-free trials. Our data shows that Response Times, overall fixations duration and fixations number are highly correlated (RT-FixNumber $r = 0.97$; RT-FixDuration $r= 0.97$; FixNumber-FixDuration $r= 0.95$). For the sake of simplicity and readability we have chosen to present here only the data about Overall Fixations Duration and Scanpaths. For scanpaths, we analysed the number of AOIs fixed by the subjects and the first fixation location. Since all the first fixations were located on the central image we decided to analyse the first fixation outside the central image AOI.

Overall fixations duration (ms ± SD)			Number of fixed AOIs (number ± SD)	
Condition	Test group	Control group	Test group	Control group
No change	1407.5 ± 370.7		2.72 ± 0.34	
Weak change	1843 ± 904.4		3.26 ± 0.74	
Non-associated change	2058.7 ± 586.3		3.24 ± 0.31	
Average	1637.16 ± 343.3	2801 ± 1042.22	3.08 ± 00.33	4.03 ± 0.67

Table 1: Overall fixations duration and numer of fixed AOIs.

First fixations location (% ± SD)			
Location	Test group inspection phase	Test group search phase	Control group search phase
Left	57 % ± 0.03	43% ± 0.16	65% ± 0.26
Centre	41 % ± 0.03	26% ± 0.18	15% ± 0.19
Right	4% ± 0.01	30% ± 20	18% ± 0.24

Table 2: First fixations location

4 Results

4.1 Information searching performances: effects of spatial coding

To verify the effects of receiving spatial information in the *inspection* phase on information searching performances in the *search* phase, we compared Eye Movements of the test group (the spatial coding group) with those of the control group (no spatial information received). Our results showed that the performance of the test group were significantly better than those of the control group. The overall duration of fixation was shorter for the test group *G1* ($t(22) = -3.67$; $p < .01$). As far as the participants' scanpaths are concerned, the results showed that the test group fixed a smaller number of AOIs before locating the target text ($t(22) = -4.35$; $p < .001$). In fact, participants in the control group browsed through an average of 4 textual AOIs before locating the target. This means that most of the time they read through all three texts (3 textual AOIs fixed) and then re-fixed their gaze on the chosen one (4th fixation). The percentage of trials in which the subject made at least one re-fixation is significantly greater for the control group: 45% versus 10% (*G1* ($t(22) = -3.40$; $p < .01$).

4.2 Information searching performances : effects of spatial and semantic distance

As far as the effects of the semantic distance on information searching performance are concerned, the results for the test group showed that information searching performances were significantly better when spatial organisation of texts didn't change (the “No change” condition) compared to when the location of the target text changed (the “Weak change” and “Non-associated change” conditions). As showed in table 1, under the “Weak change” and “Non-associated change” conditions, the number of fixations was significantly greater ($t(12)$=2.7 ; $p < .05$) as was the number of fixed textual AOIs ($t(12) = 3.76$; $p < .01$). Looking at the results in greater detail, we observe that those obtained under the “Weak change” condition (when the target text changes place with the weakly associated text) were significantly different from those obtained under the “Non-associated change” condition (when the target text changes position with the Non-associated text). In particular, pair-wise comparisons showed that overall fixation duration was significantly longer under the “Non-associated change” condition than under the “Weak change” condition ($F(1,10) = 9.32$; $p < .05$), which means that the subjects had lower performances when the Non-associated text replaced the target text.

Concerning the participants’ scanpaths, pair-wise comparisons showed that when the target text kept its original position (the “No change” condition) the average number of AOIs that the participants needed to fix with their gaze before locating the target was significantly smaller ($F(1,10) =17.0$; $p < .005$). The results presented in table 1 also showed that when the target text didn't move, the performance of the test group that had already read the texts in the inspection phase was far better than that of the control group, which had not read the texts before the search phase. Under the “Weak change” conditions, the performances of the test group approach those of the control group, and get even closer (i.e. worse) under the “Non-associated change” condition.

4.3 Information searching habits: effects of spatial cue

To conclude, we analysed first fixation location in order to detect the participants’ habits (the point from which they begin) when performing an information searching task on a multimodal document. These results are presented in table 2. They show that the participants of the test group (during the inspection phase) and participants of the control group (during the search phase) tended to begin their information searching activity from the left text, following the well-established occidental left-right reading habit (57% and 65% for the left, 41% and 15% for the centre, 4% and 18% for the right). For the test group, pair-wise comparisons showed that, during the inspection phase, the left position was fixed with the subject’s gaze significantly more often compared to the situation in the other positions *($F(1,11)= 18.6$; $p < 01$)*. On the other hand, during the search phase , the test group shifted to a more balanced distribution (43% left, 26% centre, 30% right) with no significant difference *($F(2,22)= 1.78$; $p = .19$* *) once they had explored the spatial distribution of texts during the inspection phase.

5 Discussion

The aim of this paper was twofold: i) To determine the relations between spatial coding and semantic processing. In particular we wanted to understand if semantic distance can have an effect on spatial processing and if this can also be observed in those situations in which cue and target belong to different modalities. ii) To prove that readers are able to modulate their reading habits in order to adapt their research strategies quickly to the spatial organisation of text. The results we obtained support our hypothesis.

Results presented in 4.1 show that spatial coding effects on reading and information searching can indeed be observed even in documents in which the target and the cue belong to two different modalities (e.g. textual target cued by an image). We observed that participants who could rely on spatial coding (the test group) and spatially uncued participants (the control group) each had different searching performances and scanpaths. Furthermore, we observed that the participants in the control group needed to skim all the texts and then re-fix their gaze on the right one before answering. According to our hypothesis, this result shows not only the participants had an advantage when they could attribute a spatial identity to each semantic content, but also that when they lacked this kind of spatial information (the control group), they needed to identify each text spatially before deciding on its semantics.

The results presented in 4.2 show that semantic distance has indeed an effect on spatial coding. As expected, when the target text was replaced by weak or non-associated texts, the participant’s performances worsened with the increasing of the semantic distance between the target text and the text that took its place: participants had better performances (shorter overall fixations time, fewer AOIs fixed before answering) under the Weak-change condition compared to under Non-associated change condition. We interpret these results as being caused by the co-activation of spatial and semantic information during the search activity. This activation makes it easier for the reader to process weakly associated texts because they are close to his expectations. These results are consistent with previous findings concerning co-activation of spatial and semantic information during information searching (Baccino, Pynte, 1994).

The results presented in 4.3 show that readers can easily change their reading habits and adapt their oculomotor behaviour to the local structure of information. In particular, participants in the test group easily broke the “begin from the left” rule during the search phase, when they understood that relevant information could appear with equal probability in any of the three locations. These results are consistent with previous studies about standard for web interfaces that show that users can easily adapt to unconventional structures if these structures are coherent and semantically motivated (McCarthy et al., 2003).

To resume, our experiment showed that spatial coding effects are observed even in multimodal documents, that spatial organisation is strongly intertwined with the text’s semantic organisation, and that readers easily adapt their research habits to the way information is spatially and semantically organised in a document. As far as the signalling of text organisation is concerned, this study shows that spatial organisation can be a powerful tool for signalling the semantic organisation of a document. Spatial distance can be use to signal a semantic distance and to attribute a semantic value to a location in space. Once they know were a semantic information is spatially located readers can adapt their information searching habits to the document's organisation and find information more easily. To conclude, this study allows us to make at least three recommendations to design multimodal documents such as web pages or digital documents for education effectively: *i.* Clearly link each item of content with a spatial location to obtain documents in which information can be easily searched and retrieved *ii.* Maximise the coherence between *semantic* and *spatial* distance to signal text organisation effectively *iii.* Shape readers' reading habits by adopting a coherent spatial-semantic architecture on which readers can rely when they search for information.

References

BACCINO T., PYNTE J. (1994). Spatial coding and discourse models during text reading. *Language and cognitive processes 9-2*, 143- 155.

BELLISSENS C., THÉROUANNE P., DENHIÈRE G. (2004). Deux modèles vectoriels de la mémoire sémantique: description, validation et perspective. *Le Langage et L'Homme* 39*,* 101-122.

CLARK A., CHALMERS D. (1998). The extended mind. *Analysis* 58, 10-23.

KENNEDY A. (1982). Eye Movements and Spatial Coding in Reading. *Psychological Research* 44, 313-322.

KENNEDY A. (1992). The Spatial Coding Hypothesis. In K. Rayner (Ed.) *Eye Movements and Visual Cognition*. New York: Springer-Verlag (379 - 397).

KENNEDY A., MURRAY W.S. (1987). Spatial coding and reading: Some comments on Monk (1985). *Quarterly Journal of Experimental Psychology* 39A, 649-718.

KENNEDY A., BROOKS R., FLYNN L., PROPHET C. (2003). The reader spatial code. In J. Hyönä, R. Radach, H. Deubel (Ed.) *The mind's eye: Cognitive and applied aspects of eye movement research*. Amsterdam : North-Holland. (193–212).

LANDAUER T.K., FOLTZ P.W., LAHAM D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259-284.

MCCARTHY J.D., SASSE A.M., RIEGELSBERGER J. (2004). Could I have the menu please? An eye tracking study of design conventions. Proceedings of *People and Computers XVII*, 401-414.

RICHARDSON D.C., SPIVEY M.J.(2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition* 76, 269-295.

SPIVEY M.J., GENG J.J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research* 65, 235-241.

VAN DER HEIJDEN A.H.C. (1993). The role of position in object selection in vision. *Psychological Research* 56, 44-58.

VAN DER HEIJDEN A.H.C, MUSSELER J., BRIDGEMAN B. (1999). On the Perception of Position. *Advances in Pshychology* 129, 19-37.

VIRBEL, J. SCHMID, S., CARRIO, L., DOMINGUEZ, C., WOODLEY, MPP, JACQUEMIN, C., MOJAHID, M., BACCINO, T., GARCIA-DEBANC, C. (2006). Approches cognitives de la spatialisation du langage: le cas de l'énumération. In. C. Thinus-Blanc, J. Bullier (Eds.) *Agir dans l'espace,* Paris Maison des Sciences de l'Homme (233-254).