Eye-tracking while processing written words and images: Interaction and competition between types of representation

Universidade de Lisboa
Faculdade de Letras
ONSET-CEL

Isabel Faria, Adriana Baptista, Paula Luegi, Carla Taborda

Laboratório de Psicolinguística
lab.psicoling@fl.ul.pt

This study was sponsored by the Calouste Gulbenkian Foundation

13th European Conference on Eye Movements

MTTF mean total time of fixation **MNfix** mean number of fixations
MFD mean fixation duration **I** image **C** caption **F** focus

Introduction

This study is part of a research project on eye-tracking and reading that started in our laboratory one year ago. It is also related to another major study concerning types of captions and their relationship with pictures [1] and scenes.
Our main underlying questions are as follows:
1. May an instance of written material included in an image act selectively over other internal properties of that image while perceiving it, processing it, storing it and retrieving it from memory?
2. Does the processing of written material interact with the visual processing of a scene? If so, what counts as prominent, so that it may be kept as such in our memory?
3. Do both types of representation (written and iconic) operate similarly within working memory (short term memory) and within semantic memory (long term memory)?

Methodology

Sample 20 Lisbon University students
Eyetracker ASL 504 corneal and pupil reflection, which sampled eye position at 60 Hz. Pictures were presented on a computer screen with limited observation time.

Experiment 1
Subjects were instructed to carefully observe each picture in a set of three, and at the end of each set to provide a written description of what they had seen (memory task).
The stimuli were 3 pictures: P1, P2, and P3. Two were presented without captions (P1a and P2a), three with indexical descriptive captions (P1b, P2b and P3a), and one with an additional descriptive caption (P3b).

P1

a *without caption*

b Pipe and Dutch tobacco box.

P2

a *without caption*

b Two clay oil and vinegar bottles.

P3

a Notebook, blue large carpenter pencil and glasses.

b Objects with which Mr. Manuel, a naval carpenter, would like to be berried.

Experiment 2
Memory task, similar to Experience 1. Stimuli were two real-world scenes [2].
Sets were constituted by one picture with a caption focusing on the visual focus [3] (P4a, P5a) and another picture with the caption focusing on only a sub-part of the scene, such as P4b and P5b.

P4a

P4b

a Stone fountain with two waterspouts against a railing in the S. Lázaro Garden.

b Glass and steel lamp, ideal for the ground to give bottom-up lighting.

P5a

P5b

a Oriental house, in Oporto, specializing in tea, coffee, and chocolate.

b Clothes line with some clothes hanging along the outside wall of the building.

Results

1. Interaction and recall of pictures without captions
Our eye-tracking data confirmed the existence of interaction between the written material and the image (P1a and P2a), although the written information was not necessarily brought to declarative memory in the context of recall. However, 100% recall was obtained whenever the written material was in European Portuguese (P2a).

	Image (%)	Written (%)	MTTF	MNfix	MFD
P1a	90,9	36,4	11,09	32	0,37
P2a	100,0	100,0	9,26	29	0,32

2. Interaction and recall of pictures with indexical descriptive captions
The percentage of subjects recalling written information inside pictures with indexical captions was 100% for European Portuguese written material and ≤ 50% for other languages.

	Image (%)	Written (%)	MTTF I	MNfix I	MTTF C	MNfix C
P1b	100,0	42,9	9,63	26	1,39	7
P2b	100,0	100,0	7,11	24	2,78	14
P3a	100,0	50,0	10,01	30	3,27	14

3. Recall of pictures with indexical *versus* additional descriptive captions
The inclusion of additional information in the caption provided information not contained in the picture, but did not increase the amount of retrieved written information from the image. A tendency to better recall images with indexical captions should be certified by a larger sample.

	Image (%)	Written (%)	MTTF I	MNfix I	MTTF C	MNfix C
P3a	100,0	50,0	10,01	30	3,27	14
P3b	87,5	50,0	9,40	30	3,61	17

4. Image focus and verbal focus of the caption: two areas of interest
Whenever the verbal information of the caption refers to the image focus, the time spent on the scene is always higher. One consistent finding revealed that, with this type of stimuli which contains only two areas of interest, all subjects registered a higher number of fixations in the image area than in the caption one.

	MTTF I	MNfix I	MTTF C	MNfix C
P4a	12,93	41	6,96	25
P5a	15,89	56	2,03	9

5. Captions focusing on a sub-part of the image: three areas of interest
In this situation, subjects always spent more time processing the written information of the caption than the respective visual information focused on by the caption.

	MTTF C	MNfix C	MTTF F	MNfix F
P4b	4,85	20	1,58	4
P5b	3,81	17	2,07	4

Conclusions

We obtained four main results:
- written material acting as an intrinsic property of a scene interacts with the visual processing of the scene and act selectively over other internal properties whenever the written material is presented in the first language of the subjects;
- better recall descriptions of images without captions correlated with a larger number of fixations and higher mean fixation durations;
- descriptive captions contributed to a better recall of the images. Indexical descriptive captions facilitate connections with the image for working memory and these links are sent to long term memory and recalled as such. This is also the case where captions refer to the visual focus of the image;
- captions focusing on a sub-part of the image create a third area of interest and the participants produced a higher number of fixations in the caption than in that particular area. However, in the respective memory task, the recall of the new focus was 100%.

All these results must be backed up by further data, which means broadening the sample.

References

1. Baptista, A., 2005. *Para uma Análise das Interacções entre a Legenda e a Imagem*. Dissertação de Doutoramento. Universidade de Lisboa.
2. Henderson, J.M., Ferreira, F., 2004. Scene Perception for Psycholinguists. In J. M. Henderson and F. Ferreira (Eds), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. New York: Psychology Press, pp. 2–58.
3. Rowe, Neil C., 1998. Mapping between image regions and captions concepts of depictive photographs. At http://www.nps.navy.mil/Content/CS/ncrowe/marie/aaaiwork98.html