Faria, I. H., Baptista, A., Luegi, P., Taborda, C. (2006). Interaction and competition between types of representation: An example from eye-tracking registers while processing written words and images. In José Pinto de Lima, Maria Clotilde Almeida e Bernd Sieberg (Ed.s), *Questions on the Linguistic Sign*, Lisboa: Edições Colibri e Centro de Estudos Alemães e Europeus, 115-129.

Interaction and competition between types of representation
An example from eye-tracking while processing written words and images

Isabel Hub Faria, Adriana Baptista, Paula Luegi, Carla Taborda

Laboratório de Psicolinguística, Onset-Centro de Estudos de Linguagem, Departamento de Linguística Geral e Românica, Faculdade de Letras da Universidade de Lisboa

ihfaria@fl.ul.pt

We would like to thank the Centro de Estudos Germanísticos (CEG) for the invitation to participate in this international meeting. We hope that the collaboration between CEG and our Laboratory may extend, into the future, the excellent relationship that, throughout the years, has developed among members of both institutions. For this particular occasion, allow us a special reference to José Pinto de Lima, responsible for the organisation of this meeting, and to Clotilde Almeida, colleagues and friends with whom we have shared previous work, interest and curiosity within Linguistics.

This paper is the first to refer our eyetracking research project *Registo e Análise do Movimento dos Olhos durante a Leitura* (Analysis of eye-tracking during reading) that, with the financial support of the Calouste Gulbenkian Foundation, has taken place at our Laboratory since December 2003. The project is being developed by a larger research team but, for this presentation, we are going to refer to the work carried out by the authors above.

How does the eye work?

As we all know, light rays enter the eye through the pupil and are projected upside-down on the retina. The photoreceptor nerve cells (the rods and the cones) of the retina change the light rays into electrical impulses and send them through the optic nerve to the brain, where the image is perceived.

The human visual field corresponds to the total area in which objects can be seen in the peripheral vision, while the eye is focused on a central point. Following Sanders (1993) and Rayner (1998), the visual field can be divided into three areas:

- the *foveal* area, where we have very good acuity and from where we extract the most important information about the stimulus. In this area, a stimulus can be identified without an eye movement;

- the *parafoveal* area, situated around the fixation point, from which we can still extract information about the stimulus. In this area, it is necessary to make an eye movement to identify the stimulus;

- the *peripheral* area, around the parafoveal area, from which we can only extract perceptual information (for instance, we do not extract information for reading from this area). In the peripheral area, it is necessary to make a head movement to identify the stimulus.

Eye movements are needed to locate the image at the centre of the retina. There are different types of eye movements. They may follow a moving target *(pursuit)*, move in opposite directions *(vergence)*, or occur so as to compensate head movements *(vestibular)*.

Furthermore, during the normal scanning of a visual scene, as well as during reading, eye movements are characterised by series of stops and very rapid jumps between stopping points. The stops correspond to the fixations, and the jumps to the saccades.

A *saccade* corresponds to the movement of the eye going from one fixation to another. The mean saccade size corresponds to 7-9 letter spaces. The time during which the eyes remain still is called a *fixation*. The mean duration of a fixation is usually 250 milliseconds. It is during fixations that most visual information is acquired and processed.

When reading western languages writing systems, our eyes move from left to right *(progression movement)*, but also from right to left *(regression movement)* whenever there is missing information or the reader's need to confirm something. When changing text line, our eyes sweep from the right side of the upper line to the left side of the next lower line *(return sweep)*.

International research undertaken over the last few decades has brought empirical support to the theoretical perspective that eye movements produced during reading reflect the cognitive processes that are simultaneously taking place.

Eye-tracking while processing written words and images

In this paper, our main concerns were closely related to the following questions:

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)?

Our experimental design included three sets of pictures, the first set containing 3 pictures, the second 4 and the third 3 which were to be attentively observed by each of our subjects in order for them to perform further memory tasks. An example of a used set of 4 pictures is as follows: a photograph of a ring (with caption), a photograph of a pipe (without caption), a photograph of a set of keys (with caption), a photograph of a set of stones (without caption).

At the end of each set observation, the subject was asked to recall the images just seen and to provide written descriptions of the respective pictures, referring to the largest amount of details possible related to each of them.

Does the processing of written material interact with the visual processing of a scene?

Our eye-tracking data and also some of the recalled written descriptions mainly confirm the existence of interaction between written material and the images inside a picture scene, such as in Figure 1, where one can easily observe that the tobacco box contains written information, as well as the presence of a letter ‘C’ inscribed on the pipe.

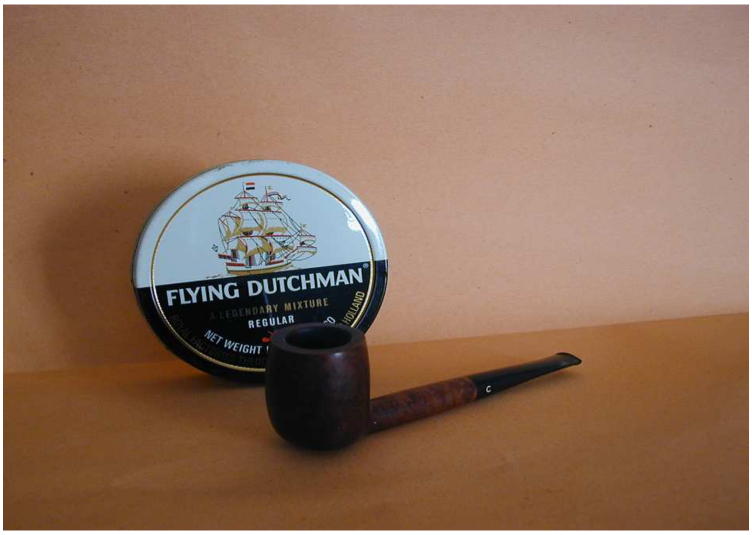

Figure 1

The written material seems to be taken as an intrinsic property of an object or of the scene itself, although not necessarily brought to declarative memory whenever the subjects were asked to recall and describe what they had seen. One possibility for this apparent lack of attention may be that this picture's internal written material is not really processed as such. This could happen, whenever the written material belongs to a language that the reader does not know.

For example, when recalling the pipe picture that was shown without caption, we receive really contrasting descriptions such as (1) (a) and (b):

(1) (a) "*Cachimbo castanho*"
(Brown pipe)
(SMV: Total time of fixation - 2,249)

(b) "*Cachimbo e caixa de tabaco. O cachimbo tinha a ponta (parte que se põe na boca) de cor preta com a inscrição da letra 'C' na parte oposta, sendo o restante cachimbo de madeira castanha escura. A caixa de tabaco era redonda e tinha a tampa virada para a frente, tinha os transbordos pintados a dourado. A parte de cima da tampa era branca com o desenho de uma caravela enquanto a parte de baixo era preta com indicações relativas à origem do tabaco e as suas qualidades que se encontravam um pouco escondidas pelo cachimbo*"
(Pipe and tobacco box. The pipe has a black tip (the part that one puts in the mouth) with a letter 'C' inscribed on the opposite side, the rest of the pipe being dark brown wood. The tobacco box was round and had the lid facing the front, with painted gold edges. The upper part was white with a picture of a *caravel*, while the lower part was black with a small section of the information about the origin of the tobacco and its qualities hidden by the pipe)
(JMS: Total time of fixation - 5,587)

Equivalent contrasts can be easily observed when we look at the respective eye-tracking registers. Description (a) was provided by the subject whose eye movements are represented in Figure 2, and description (b) was produced by the subject whose eye-tracking is represented in Figure 3.

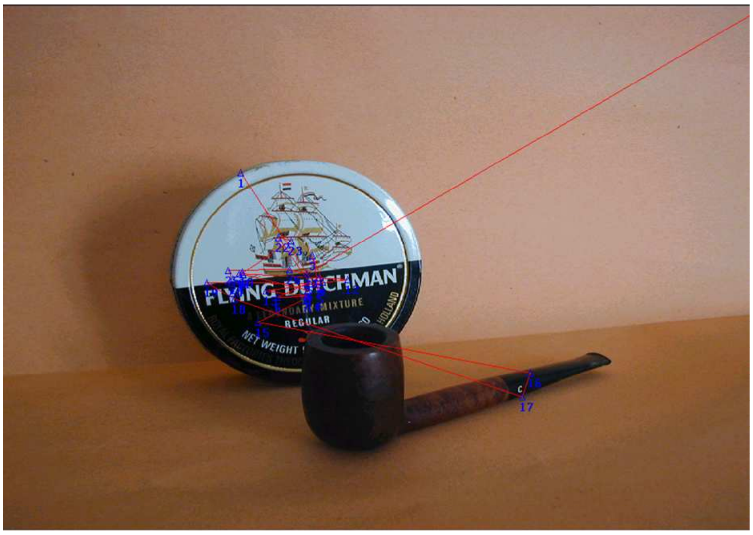

Figure 2 – Description (a) – SMV: Total time of fixation - 2,249.

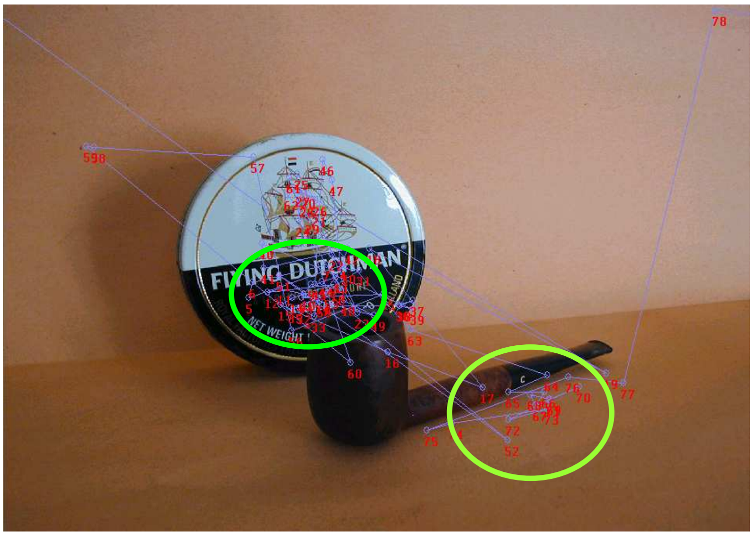

Figure 3 – Description (b) – JMS: Total time of fixation - 5,587.

Together with the larger number of fixations by the subject who gave the recalled description (b), we may observe a concentration of sub-sets of fixations around the areas containing written material (the top of the tobacco box and the letter ‘c’ inscribed in the black part of the pipe).

We therefore proceeded to find out whether this tendency (a higher number of total fixation time being at the basis of longer and highly detailed recalled descriptions) would be sustained throughout the other stages of our experiment.

Is there competition between iconic and written signs?

Using a simple picture, without a caption, of a set of pumice stones (as in Figure 4), we were able to confirm the previous tendency. In fact, more complex recalled descriptions of images without caption correlate to a larger fixation time, as a whole.

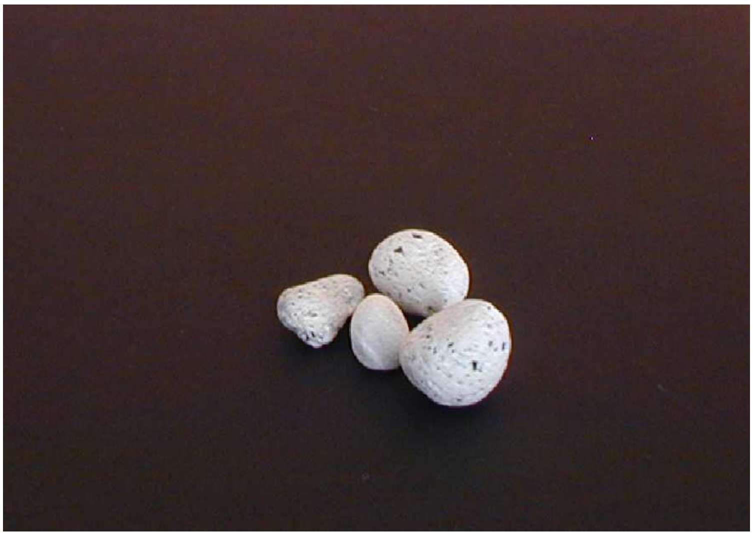

Figure 4

This could easily be found, merely by noticing the increasing complexity in answers in (2), from (a) to (d):

(2) (a) *"Pedrinhas pequenas, brancas"*
(Small white stones)
(SMV: Total time of fixation - 0,617) (Figure 5)

(b) *"Quatro pedras brancas"*
(Four white stones)
(ARP: Total time of fixation - 1,852)

(c) *"Conjunto de pedras brancas. Quatro pedras brancas: uma pequenina (no centro), uma um pouco maior e duas de tamanho relativamente grande (agrupadas em forma de semi-círculo em redor da mais pequenina)"*

(Set of white stones. Four white stones: a little one (in the centre), a slightly larger one, and two relatively big ones (grouped in a semi-circle shape around the smallest)

(JMS: Total time of fixation - 2,001)

(d) *“Quatro pedras brancas, 2 de dimensão maior, as outras 2 eram mais pequenas. Em três existiam alguns buracos, que pareciam pintas pretas. Na pedra mais pequena não existia nenhum buraco, era toda branca.”*

(Four white stones, 2 were bigger while the other 2 were smaller. In three of them they also had some holes which looked like black spots. The smallest stone, which was completely white, had no hole in it).

(MSS: Total time of fixation - 8,669) (Figure 6)

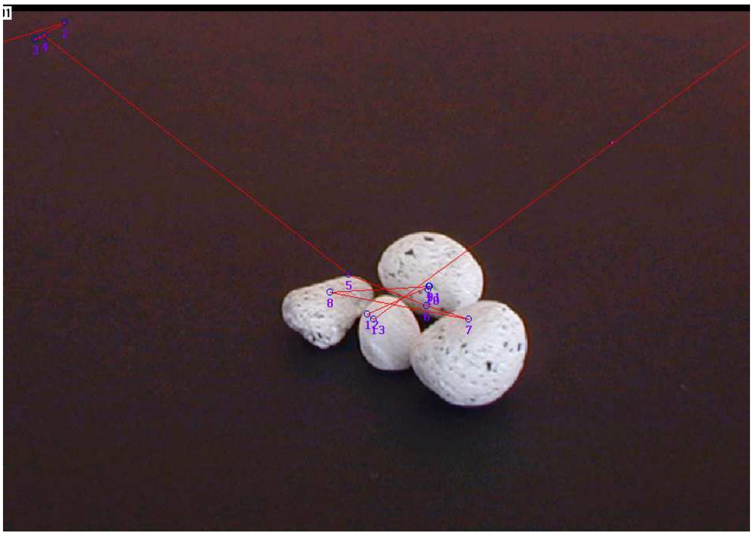

Figure 5 – Description (2a) – SMV - Total time of fixation: 0,617.

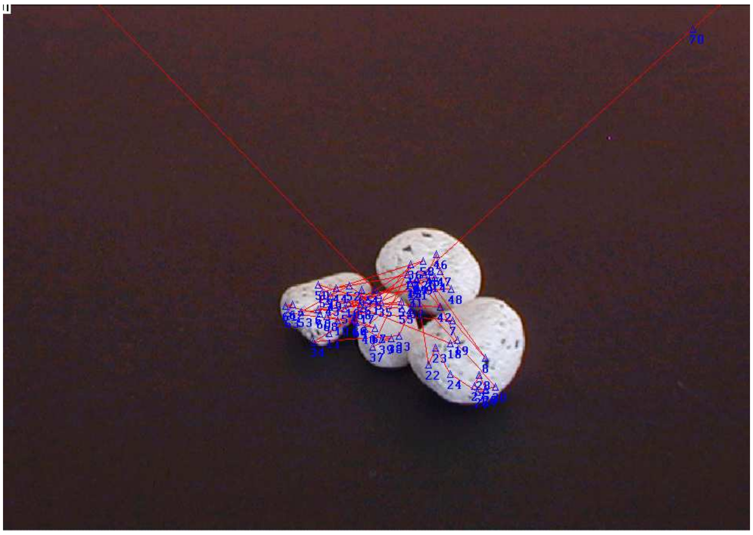

Figure 6 – Description (2d) – MSS - Total time of fixation: 8,669.

In order to learn whether captions help in facilitating the process of classifying and attributing meaning to a scene or to parts of a scene, we used the same picture, this time containing a descriptive caption, both indexical and additive (A set of four small, well-worn, very smooth pumice stones used in washing laundry (stonewash) to give it its worn look) as shown in Figure 7, to a different group of subjects. Our main concern was to observe whether the integrated meaning of a caption would allow for an easier retrieval from memory of what was visually represented. By ‘caption’ we mean a text instance which, situated at the periphery of a reproduced image, establishes a factorial relation with this image so that it produces particular meanings and particular deictic relations. The indexical nature of a caption allows for attention to be focussed on a given dimension of that image. The additive nature of a caption recruits further attention on elements which are not directly represented in the image.

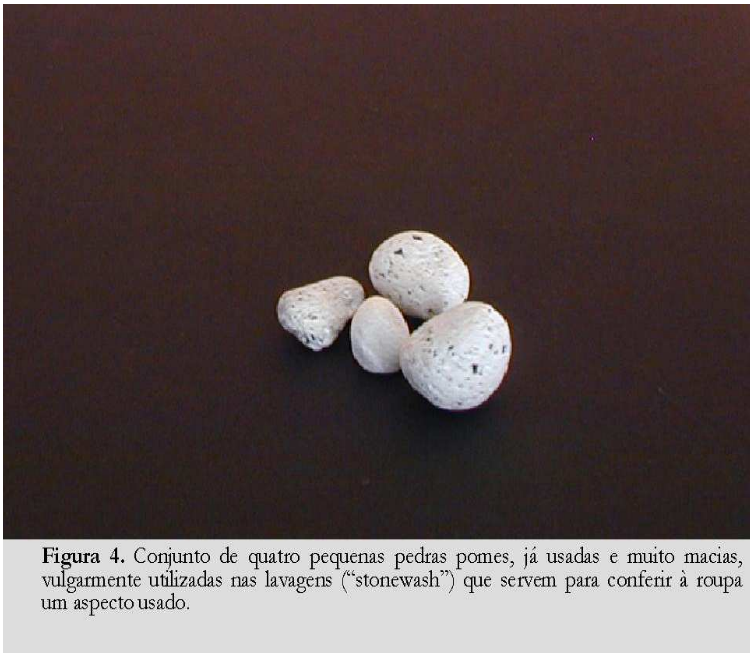

Figura 4. Conjunto de quatro pequenas pedras pomes, já usadas e muito macias, vulgarmente utilizadas nas lavagens (“stonewash”) que servem para conferir à roupa um aspecto usado.

Figure 7 – Set of four small well-worn, very smooth pumice stones used in stonewashing laundry to give it its worn look.

The first interesting result was that all the recalled descriptions provided the category name ('pedras pomes', 'pumice stones') introduced by the caption, even when dealing with problems concerning the plural grammatical form of the compound, as we may see by the examples in (3):

(3) (a) *"Pedra pomes brancas que são utilizadas pelos humanos em certas ocasiões"*
(White pumice stones which are used by humans on certain occasions)
(ACG: Total time of fixation - 2,818)

(b) *"Quatro pedras pomes, duas delas eram relativamente grandes comparando com as outras, sendo cada uma delas um pouco mais pequena. Pedras essas que tinham umas pintas pretas."*
(Four pumice stones, two of them relatively large when compared with the others, that were each a little smaller than the other. The stones had some black spots on them)
(SSS: Total time of fixation - 6,021)

(c) *"4 pedra-pomes usadas e macias; usadas para gastar roupa; ('stone wash')"*
(4 soft, well-worn pumice stones for stone-washing clothes
(RMR: Total time of fixation - 5,783; 21 fixations in the image, with TTF - 2,733; 30 fixations in the caption, with TTF - 3,05)

These results empirically allow us to confirm that the used caption, with its simultaneous indexical-additive nature, contributed to successful categorisation, facilitating the process of classifying and attributing meaning to the scene, by introducing 'names' (descriptions) that helped the subjects to adequately categorise the stones as 'pumice stones'.

Nevertheless, we may ask whether the integrated meaning of a caption allowed for an easier retrieval from memory of what was visually represented. If so, how did it operate? Registering our subject's RMR eyes when reading the caption, we found that his largest single fixation was positioned at the word 'pomes' (pumice), and that several fixations, after regressions, on the word ‘stonewash’ presented a TTF of 0,617. These two factors seemed to account for the accuracy of RMR’s description (Figure 8).

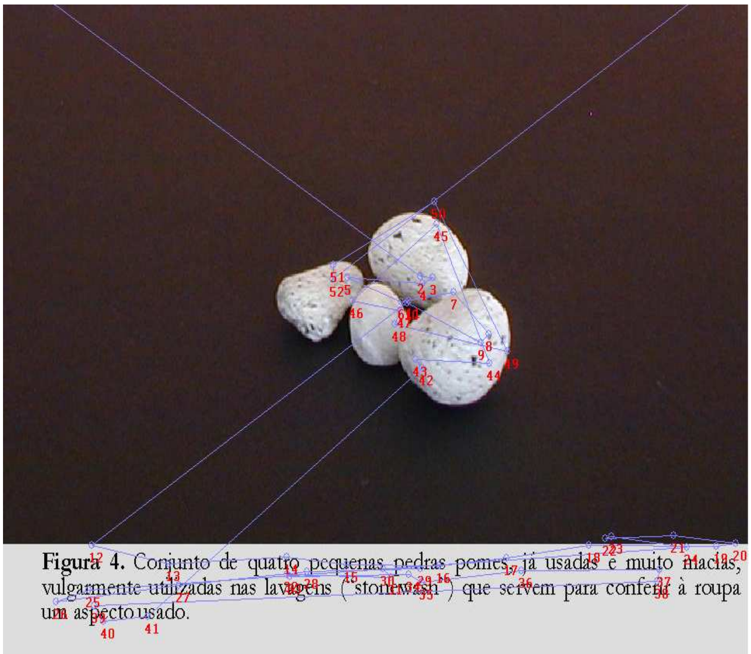

Figura 4. Conjunto de quatro pequenas pedras pomes, já usadas e muito macias, vulgarmente utilizadas nas lavagens ("stonewash") que servem para conferir à roupa um aspecto usado.

Figure 8 - Description (3c) – RMR

Is there competition between iconic and written signs?

Does a caption compete with the information contained in the respective image?

By comparing some of the recalled descriptions provided of the same image, with and without caption, we were able to find obvious similar recalls which contained information that was not included in the written information of the provided caption. This seemed to point to the fact that, although successfully introducing new information such as the category name, the caption did not cover, and therefore did not control, the visual information of the scene that was relevant or salient for the subject.

Examples of similar recall descriptions:

(4) (a) Image with caption

“Quatro pedras pomes, duas delas eram relativamente grandes comparando com as outras, sendo cada uma delas um pouco mais pequena. Pedras essas que tinham umas pintas pretas.”

(Four pumice stones, two of them were relatively large when compared with the others, that were each a little smaller than the other. The stones had some black spots on them)
(SSS: Total time of fixation - 6,021; in the image - 3,070; in the caption - 2,951)

(b) Image without caption
“Quatro pedras brancas, 2 de dimensão maior, as outras 2 eram mais pequenas. Em três existiam alguns buracos, que pareciam pintas pretas. Na pedra mais pequena não existia nenhum buraco, era toda branca.”
(Four white stones, two were bigger while the other two were smaller. There were some holes in three of them which looked like black spots. The smallest stone, which was completely white, had no hole in it).
(MSS: Total time of fixation - 8,669)

Does the visual information resist and somehow overtake the written information of a caption?

Although contributing explicitly to categorization and naming, we verified that the used captions did not prevent the subjects from accessing other information visually contained in the picture or judged by the subject as implicit. A certain unexpected ‘control free’ nature of the used captions could provide a possible explanation for obtaining similar recalls for a picture from different subjects, seen either with or without a caption. Those observations lead to the formulation of new hypothesis, and to further research.

Initially, we had drawn up captions that were verbally focused, i.e., referred to the visual focus of the whole image. Following Rowe (1994), on images with pedagogic aims, the visual focus corresponds to sets of areas which must contain the following five characteristics: it should be large; have its gravity centre near the centre of the picture; not touch the picture frames; have sharp outlines, be significantly differentiated from other non-focus areas of the picture.

In order to question the strength of the visual information of a picture, we elaborated new captions, this time verbally focusing on a part of the scene distinct from the general visual focus.

Therefore, for this new experiment, we were dealing with two captions for the same scene. In one case, we had a caption verbally focusing on the scene's visual focus, so the subject had to deal with two areas of interest: the caption and the whole image (as in Figure 9).

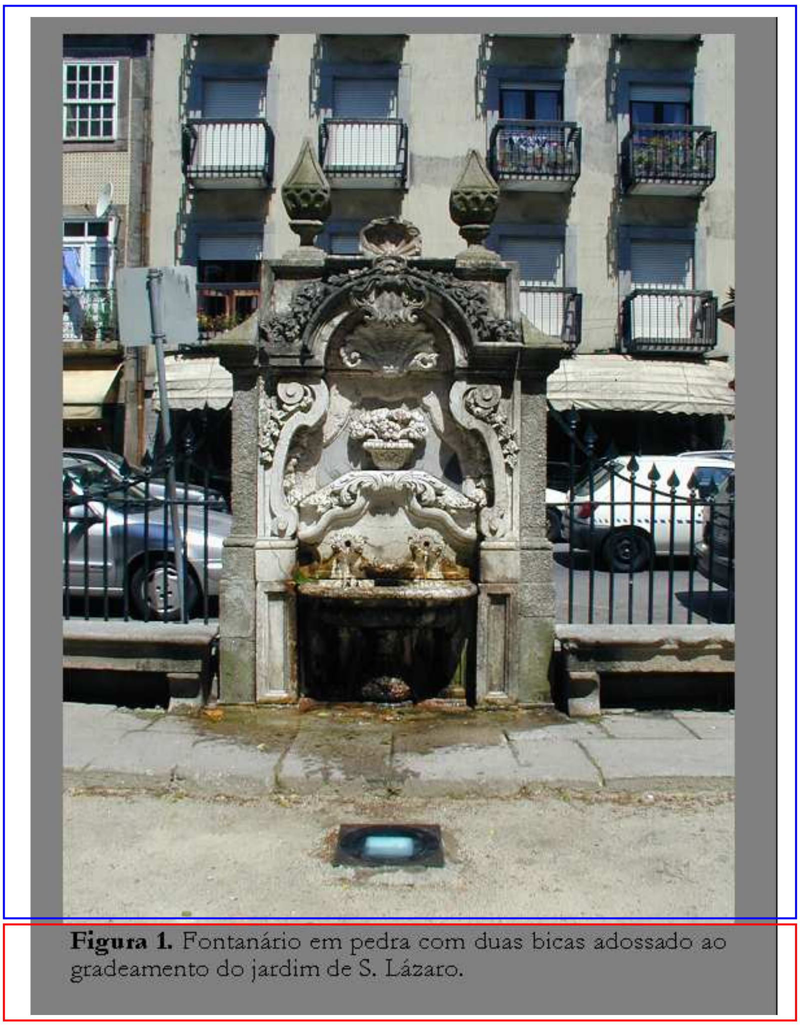

Figura 1. Fontanário em pedra com duas bicas adossado ao gradeamento do jardim de S. Lázaro.

Figure 9 – Stone fountain with two waterspouts against a railing in the S. Lázaro Garden.

In the other case, the caption verbally referred to a sub-part of the image, so the subject had to deal with three areas of interest: - the caption, the sub-part of the scene verbally focused on by the caption, and the rest of the image (as in Figure 10).

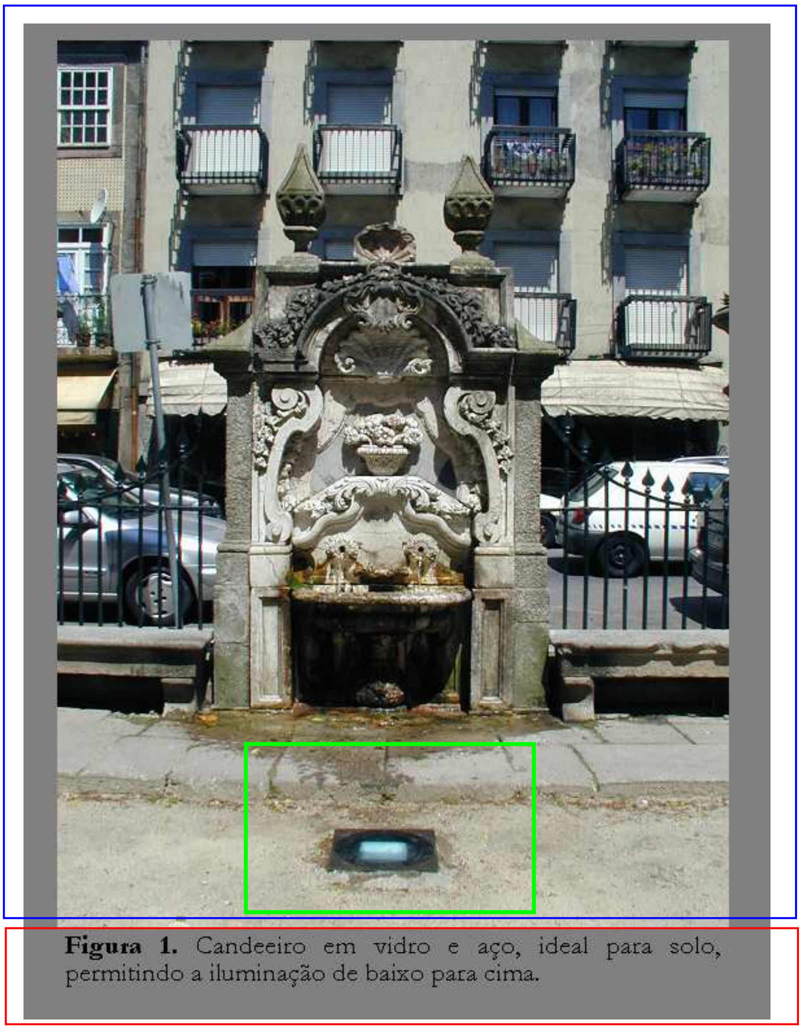

Figura 1. Candeeiro em vidro e aço, ideal para solo, permitindo a iluminação de baixo para cima.

Figure 10 – Glass and steel lamp, ideal for the ground to give bottom to top lighting.

The recalls obtained for the first type of stimulus (Verbal Focus = Visual Focus) followed the previous pattern found for recalling images with general descriptive captions, mainly containing the category name, an optional description and reference to some visual details. As in example (5):

(5) *"Fontanário em pedra, com duas bicas no Jardim de S. Lázaro. Atrás encontra-se um prédio, com 8 janelas, 3 das quais estão abertas. Ao lado há outra casa, que tem 2 janelas. Ambas estão abertas. Entre a fonte e os prédios existe uma estrada, onde se encontram carros. Há um sinal de trânsito que está virado de trás para a frente"*

(Stone fountain with two water spouts in the S. Lázaro Garden. Behind it there is a building with 8 windows, 3 of which are open. Beside it, there is another house

which has 2 windows. Both are open. Between the fountain and the buildings is a road where we find some cars. There is a traffic sign which is turned back to front)
(AMP: ¾ of the fixations in the image; ¼ in the caption)

One consistent finding revealed that, with this type of stimulus, containing only two areas of interest, all subjects registered a higher number of fixations in the image than in the caption.

However, the recalls of the second type of stimulus, containing three areas of interest, the third being put into focus by the caption, did not follow the previous pattern found for type one stimulus. 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. Time spent processing the rest of the visual information varied with the subject. For instance, the subject SMV registered 22 fixations in the caption, but only 4 in the verbally restricted focus area, and 9 fixations for the rest of the scene.

Furthermore, a consistent finding was observed in this situation: for all subjects, the number of fixations in the caption was always higher than the number of fixations in the visual focus. But this, as well as other findings should, of course, be backed-up in the near future by studying a larger group of subjects.

References

Baptista, A. (2005) *Para uma Análise das Interacções entre a Legenda e a Imagem*. Dissertação de Doutoramento. Universidade de Lisboa (a aguardar marcação de provas).

Castro Caldas, A. (1999) *A Herança de Franz Joseph Gall: O cérebro ao serviço do comportamento humano*. Amadora: McGraw-Hill.

Chalupa, L. M., Werner, J. S., (eds), (2003) *The Visual Neurosciences*. Cambridge: Mass. MIT Press.

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.

Rayner, K. (1998) *Eye Movements in Reading and Information Processing: 20 Years of Research*. Psychological Bulletin, n.º 3, vol. 124, pp. 372–422.

Rayner, K., Liversedge, S. P. (2004) Visual and Linguistic Processing During Eye Fixations in Reading. In J.M. Henderson and F. Ferreira (eds.), *The Interface of Language, Vision, and Action*. New York: Psychology Press, pp. 59–104.

Rowe, Neil C. (1994) Inferring depictions in natural language captions for efficient access to picture data. At http://www.cs.nps.navy.mil/research/marie/indcap.html

Sanders, A. E (1993) Processing information in the functional visual field. In G. d'Ydewalle & J. Van Rensbergen (eds.), *Perception and cognition: Advances in eye movement research*. Amsterdam: North Holland, pp. 3–22.