

Luegi, P., Costa, A., Faria, I. H. (2010). Eye movements during reading. In Leonor Scliar-Cabral (Ed.), *Psycholinguistics Scientific and technological challenges, Selected papers: 8<sup>th</sup> International Congress of the International Society of Applied Psycholinguistics*, EDIPUCRS – Editora Universitária da PUCRS, Porto Alegre, p. 159-172.

*Paula Luegi*  
*Maria Armanda Costa*  
*Isabel Hub Faria*

Laboratório de Psicolinguística, Centro de Linguística  
da Universidade de Lisboa (CLUL)

---

## **Eye movements during reading<sup>1</sup>**

### **1. Introduction**

In this paper we present the results of a cluster analysis of data from eye-movement during text reading<sup>2</sup>. Two texts with different topics were adopted and two versions of each text were created: one version, corresponding to the original text, acted as a control text; in the second text, syntactic changes in some sentences were introduced. The texts were manipulated and first used by Armanda Costa in 1991 in which she analysed texts read aloud. The aim of the study was to identify the strategies used by the readers when they experienced difficulties in processing texts with different topics and with syntactic manipulations.

In the present study, as in Costa (1991), it is taken for granted that understanding lexis, classifying words according to syntactic categories, integrating these words into syntactic structures and all the other processes necessary to interpret a sentence are undertaken immediately, quickly and nearly always efficiently (COSTA, 2005). According to LaBerge and Samuels (1985), some of these processes or stages are undertaken automatically, thus allowing the reader's attention to be directed to other matters: attention ceases to focus so closely on de-codifying or on the specific mechanisms of reading itself and starts taking into consideration the meaning of the sentence. Nevertheless, there are situations where the automatic response breaks down and the reader is obliged to reformulate or reanalyse his/her initial interpretation. These situations happen mostly when there are complex, ungrammatical or even temporarily ambiguous structures.

Briefly, reading is considered to be an incremental process where the reader interprets the linguistic material (morphemes, words, phrases) as soon as it is found, and so, any difficulty experienced during information processing will have immediate impact. It is assumed in psycholinguistics research that some experimental paradigms can detect and reflect whether the reader has experienced any difficulty during reading. There are several indicators that can be registered and analysed to identify

---

<sup>1</sup> This study was conducted by Paula Luegi for her Master's thesis under the supervision of Professors Isabel Hub Faria and Maria Armanda Costa. The eye-tracking equipment used in this study was financed by the Calouste Gulbenkian Foundation.

<sup>2</sup> The results of other analysis may be seen in Luegi (2006) or in Luegi, Costa and Faria (2007).

the difficulties experienced by readers: Costa (1991) for example, studied prosodic variables in reading aloud, such as the duration and locality of pauses, the presence and the duration of hesitations and repetitions, the speed of reading, among other factors; Luegi (2006) studied silent reading, analysing the number and the duration of fixations, the time taken to read the texts, sentences or excerpts of some sentences, the existence of regressions, among other factors.

In order to analyse the different variables which are taken to be more or less direct indicators of the cognitive mechanisms involved in language processing, it is necessary to rely on experimental paradigms that allow us to register (and analyse) these variables. In the experiment we conducted, we used eye-tracking during reading. This technique is based on the assumption that by studying eye movements, we are able to perceive the cognitive processes involved in understanding linguistic information at the moment it is happening (STARR; RAYNER, 2001, p. 156). So, we will begin this paper giving a brief description of eye-movements behaviour during reading.

## **2. Eye movements during reading and text processing**

In 1878/79, Émile Javal, a French ophthalmologist, noted that, during reading, our eyes do not move in a linear way from the beginning to the end of the line but, rather, they move forward in short, very fast jumps that are punctuated by brief pauses. The jumps are called saccades while the stops are called fixations.

When reading, the saccades are called progressive when they move from left to right in Western languages and regressive when they happen in the opposite direction. Regressive movements represent about 15% of all saccades in reading and occur mainly when the reader experiences difficulties during information processing. Due to a phenomenon called saccade suppression (MATIN, 1974), we are not able to perceive whether anything has happened in the visual stimulus when moving the eyes, that means, during a saccade. Saccades have an average length of 7-9 letter-spaces (the characters plus the spaces between the words).

Only during the fixations (new) information is collected and, owing to this, several fixations are made when we are regarding a stimulus. Since the eye has a limited acuity, during a fixation we can only manage to collect information covering a small region: 17-19 letter-spaces. Therefore during a fixation on a line of text, we collect information from the left and from the right of the fixation point. During the reading of Western languages, more information is collected from the right than from the left of the point of fixation and the inverse happens when reading languages that are written and read in the opposite direction. Therefore, the visual field in reading is asymmetrical, with 3-4 letter-spaces placed to the left of the fixation point, and 14-15 to the right. Apart from collecting useful linguistic information for processing from a space 17-19 characters long, called the parafoveal region (covering both foveal and parafoveal vision), we only really extract useful information for word processing

from a space 9-12 characters long, called the foveal region (3-4 letter-spaces to the left and 6-8 letter-spaces to the right).

In reading, the mean fixation duration is considered to be about 250 milliseconds (hereafter referred to as ms). Nevertheless, this measurement is fairly variable. On the one hand, readers with different proficiencies in reading have different fixation durations (more fluent readers are expected to have lower fixation times); on the other hand, the same reader has variable fixation durations depending upon the difficulty of the text being read. Apart from reading proficiency and the overall complexity of the texts, other linguistic factors may influence the duration and the frequency of fixations, such as the frequency of the word that is being fixated, its length and the context in which it appears, among other factors. The syntactic complexity of structures may also be responsible for increasing the duration and the frequency of fixations. However, the influence exerted by these factors is not only reflected in the duration and frequency of the fixations, but also in the direction and length of the saccades, with a likely increase in regressions and in shorter saccades.

It is precisely the differences in eye-movement patterns, which are closely connected to the linguistic characteristics of the stimulus being processed, that make this method so interesting to better understand the way the brain works during language processing. It is assumed that in more complex reading situations, where cognitive processing is overcharged, there is a change in eye-movement patterns that reflect the cognitive processes involved in language processing. Eye movements, therefore, act as more or less direct and valid indicators of the way the brain works above all when the reader has to deal with more complex situations involving information processing of the written language.

### 3. Experiment

The aim of the statistical analysis we present in this paper is to make an overall assessment about whether:

- (i) Eye movements are sensitive to the syntactic structure of the sentence which is being read.
- (ii) Syntactic complexity causes changes in eye-movement patterns.

More particularly, this study wishes to analyse whether ungrammatical or ambiguous syntactic structures with common properties cause eye-movement behaviours that are similar to one another. Likewise, it is also to see whether grammatical constructions having different properties cause distinct eye-movement patterns. As we shall describe shortly, we contrasted two large groups of constructions: structures involving constituent's movement and structures without constituent's movement, checking to see whether the eye movements registered during reading come within proximity to each other or not.

### 3.1 *Stimuli*

As stimuli we used two texts that had been previously used by Costa (1991). In what concerns the layouts, the two texts have a very similar textual structure in that both have a title which introduces the topic, they have identical structural information in terms of the distribution of the topics or paragraph organisation, and both texts have a concluding paragraph that is introduced by a rhetorical question. Both texts have approximately the same number of words. Likewise, at the level of linguistic structure, both texts are fairly similar, with the sentence syntactic structures of the same type along the texts.

Despite these similarities, the two texts differ as to the familiarity of their topics. One text is about a well-known topic, *Campo de Ourique* (T1), while the other is about a scientific topic, *O Isolamento termo-acústico* (T2), distant from the reader's universe of reference and, consequently, with less well-known words. This manipulation introduces the independent variable TOPIC.

Two versions of each of the texts were made: a control version (T) and a version in which some syntactical structures were manipulated (T'). This manipulation gave rise to an independent variable, Deterioration of the Syntactic Level (henceforth DSL). The target syntactic structures were chosen on the basis of their grammatical properties in European Portuguese (EP from now on) and they were duly inserted into contexts.

In each of the texts contexts of analysis were singled out. They will be described below (the contexts under study are written in italics and the vertical bars (|) indicate the different regions of analysis: A, B or C):

Context 1 (C1): the clitic was placed in post-verbal position in a relative clause, setting up an ungrammatical situation in EP.

- T1 – o bairro colorido e calmo, | *que se vislumbra* |<sub>A</sub> *através dos eléctricos em movimento* |<sub>B</sub>  
 the colourful quiet quarter, | *which (clitic) is glimpsed* |<sub>A</sub> *at through the moving trolley-car* |<sub>B</sub>
- T1' – o bairro colorido e calmo, | *que vislumbra-se* |<sub>A</sub> *através dos eléctricos em movimento* |<sub>B</sub>  
 the colourful quiet quarter, | *which is glimpsed-(clitic)* |<sub>A</sub> *at through the moving trolley-car* |<sub>B</sub>
- T2 – os múltiplos sons de choque, | *que se captam* |<sub>A</sub> *no interior de cada edifício* |<sub>B</sub>  
 the multiple sounds of crashes, | *which (clitic) are captured* | *inside each building* |<sub>B</sub>
- T2' – os múltiplos sons de choque, | *que captam-se* |<sub>A</sub> *no interior de cada edifício* |<sub>B</sub>  
 the multiple sounds of crashes, | *which are captured-(clitic)* | *inside each building* |<sub>B</sub>

Context 2 (C2): the clitic complement of the verb which acts as a direct object is omitted, giving rise to an ungrammatical sentence.

- T1 – A vida deste bairro mundano [...] | *revela-se* |<sub>A</sub> *ao virar de cada esquina das suas ruas de passeios largos.* |<sub>B</sub> *O casario heterogéneo do bairro* |<sub>C</sub>  
 The life of this mundane quarter [...] | *reveales itself*<sub>(clitic)</sub> |<sub>A</sub> *at the turn of each corner of it streets with their wide pavements.* |<sub>B</sub> *The quarter's heterogenous rows of houses* |<sub>C</sub>
- T1' – A vida deste bairro mundano [...] | *revela* |<sub>A</sub> *ao virar de cada esquina das suas ruas de passeios largos.* |<sub>B</sub> *O casario heterogéneo do bairro* |<sub>C</sub>  
 The life of this mundane quarter [...] | *reveales* |<sub>A</sub> *at the turn of each corner of it streets with their wide pavements.* |<sub>B</sub> *The quarter's heterogenous rows of houses* |<sub>C</sub>

- T2 – A resolução deste problema [...] | *centra-se* |<sub>A</sub> *na existência de meios técnicos actuanes na oposição à propagação de ruídos.* |<sub>B</sub> *Uma das soluções mais eficazes* |<sub>C</sub>  
 The solution of this problem [...] | *centers itself*<sub>(clitic)</sub> |<sub>A</sub> *on the existence of technical means countering the propagation of noise.* |<sub>B</sub> *One of the most efficient solutions*|<sub>C</sub>
- T2' – A resolução deste problema [...] | *centra* |<sub>A</sub> *na existência de meios técnicos actuanes na oposição à propagação de ruídos.* |<sub>B</sub> *Uma das soluções mais eficazes* |<sub>C</sub>  
 The solution of this problem [...] | *centers* |<sub>A</sub> *on the existence of technical means countering the propagation of noise.* |<sub>B</sub> *One of the most efficient solutions* |<sub>C</sub>

Context 3 (C3): where the Subject is in a post-verbal position in a non-marked declarative clause, giving rise to temporary ambiguity because the NP to the right of the transitive Verb may be interpreted as a Subject or as an Object in a null Subject construction<sup>3</sup>.

- T1 – | *as donas de casa atarefadas procuram os melhores produtos frescos* |<sub>A</sub>  
 | *busy housewives seek the best fresh produce* |<sub>A</sub>
- T1' – | *procuram as donas de casa atarefadas os melhores produtos frescos* |<sub>A</sub>  
 | *seek busy housewives the best fresh produce* |<sub>A</sub>
- T2 – | *o painel ISOLPAN apresenta vantagens excepcionais* |<sub>A</sub>  
 | *the ISOLPAN panels display exceptional advantages* |<sub>A</sub>
- T2'<sup>4</sup> – | *apresentam os painéis ISOLPAN vantagens excepcionais* |<sub>A</sub>  
 | *display the ISOLPAN panel exceptional advantages* |<sub>A</sub>

Context 4 (C4): where the Verb is in final position in a WH-question with simple WH-morpheme, thus resulting in an ungrammatical structure.

- T1 – | *Como reagem os moradores de Campo de Ourique?* |<sub>A</sub> *Receiam que as vizinhas torres do progresso* |<sub>B</sub>  
 | *How do react the Campo de Ourique residents?* |<sub>A</sub> *They fear that the neighbouring towers of progress* |<sub>B</sub>
- T1' – | *Como os moradores de Campo de Ourique reagem?* |<sub>A</sub> *Receiam que as vizinhas torres do progresso* |<sub>B</sub>  
 | *How do the Campo de Ourique residents react?* |<sub>A</sub> *They fear that the neighbouring towers of progress* |<sub>B</sub>
- T2 – | *Como actuam os especialistas em isolamento?* |<sub>A</sub> *Defendem que a eleição de materiais e de técnicas* |<sub>B</sub>  
 | *How do actuate the insulation specialists?* |<sub>A</sub> *They argue that choosing the materials and the techniques* |<sub>B</sub>
- T2' – | *Como os especialistas em isolamento actuam?* *Refendem que a eleição de materiais e de técnicas* |<sub>B</sub>  
 | *How do the insulation specialists actuate?* |<sub>A</sub> *They argue that choosing the materials and the techniques* |<sub>B</sub>

<sup>3</sup> Since EP is a Null Subject Language, the manipulated sentence could be interpreted as having a null subject. So, this sentence could have three interpretations:  $\emptyset_{\text{Null Subject}} \text{ V O}$ ;  $\text{V S O}$ ;  $\text{V O S}$ .

<sup>4</sup> In this structure, the number of the first NP has been changed (*the panel* has become *the panels*). We decided to change this so as to set up an ambiguous situation where it is necessary that both the post-verbal NPs agree with the verbal inflection in both gender and number.

In the set of four altered sentences, there are two distinct situations: i) ungrammatical constructions without a direct bearing on interpretation: a clitic in an enclitic position where proclisis is mandatory, and a Subject in a pre-verbal position in a WH-question; and ii) misleading constructions making the reader hesitate before interpreting them, such as, omitting the complement of the verb which may occupy a final position (after the prepositional phrase), or the post-position of the Subject in a declarative sentence where it may be interpreted as an Object in null-subject construction or, correctly, as a post-poned Subject. These constructions, which give rise to disturbances in interpretation, may demand distinct processing efforts that may be visible in eye movement behaviour.

Within each of the contexts, different regions of analysis are established (the values of the contexts are the sum of the values of the regions), indicated by the bars (|) as seen above. These regions have been set up since it is predicted that, firstly, the ungrammatical or ambiguous structures are not always immediately identified in the region in which they occur, but may only be noticed in the following words. Secondly, even if they are detected immediately, some effects of the detection may occur afterwards. In other words, a phenomenon known as the spill-over effect<sup>5</sup> may occur. Each one of the contexts has been divided into two regions of analysis: Region A and Region B. Region A is the critical region and covers the places where the problem has been set (however, it may not always be detected there – as referred above). Region B is called the post-critical region which covers the words coming after the problem; this is the region where it is hoped that, as extra material is given, the problem will be confirmed and solved, or from where, it is hoped, regressions will be made to the region where the problem originated.

### ***3.2 Participants***

The sample was composed by 20 university students or recent graduates. All the participants were native speakers of European Portuguese and their average age was 22 years and 9 months.

The 20 participants were divided into two subgroups: 10 participants read T1 without DSL and T2 with DSL; the other 10 participants read T1 with DSL and T2 without DSL. To avoid effects of order, reading was done alternately: half of the participants of each group read T1 first while the remaining participants read T2 first.

### ***3.3 Procedure***

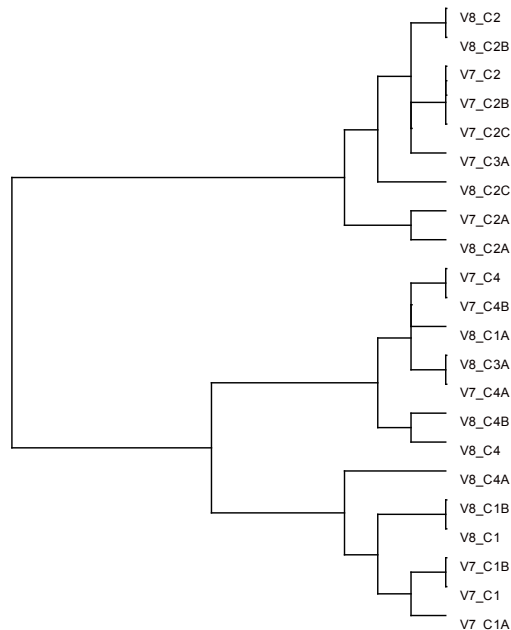
The texts were divided into three parts where each part comprised two paragraphs that were shown on different PowerPoint slides and presented on a computer screen.

---

<sup>5</sup> Reading time on a word may be increased by difficulties in the processing of the word before, indicating that the processing of that (complex) word was not finished when it was abandoned and a new word was fixated.

As the texts were being read, the participants' eye movements were registered by an ASL 504 Model working at a recording speed of 60 Hz. This is a remote-controlled system comprising a camera, placed under the screen showing the texts at a distance of 60cm from the participant, and that emits an ultraviolet ray of light which creates two reflection points: one on the cornea and the other on the pupil. The position of the eye is calculated on the basis of the distance between these two points (when the eye moves, the distance between the reflexion of the pupil and the reflexion of the cornea changes). To avoid head's movement, we used a chin rest.

Before the experiment was started, the participants were told about the procedure. They were told that they were going to read the texts shown on the screen and that, since they had to read and understand the text, at the end of each text they would answer some questions about what they had just read. After the instructions had been read and explained in detail to the participants, the equipment was calibrated<sup>6</sup> and then, in order to check whether it was working properly, participants read a small sample text. If needed, the equipment was adjusted or if it was satisfactory, the task was begun.



**Figure 1:** Dendrogram Campo de Ourique

<sup>6</sup> So that the correct position of the eye of each participant is obtained, it is necessary to calibrate the system at the beginning of the test. To do this, the participant has to look at 9 numbered points on the screen (which coordinates are pre-defined). The eye's position is then registered when it fixes upon each of these points. Afterwards, the system calculates (according to the vertical and horizontal coordinates), the position of the eye and its movements on the basis of these key-points.



### 3.4 Cluster analyses results

Owing to the lack of space, we will present only the results of a multivariate data analysis, in this case, the Cluster Analysis (for more details, consult Luegi (2006)). The Cluster Analysis allows us to group into classes the conditions that resemble each other the most and that consequently are closer to or related to one another (the result is presented on a dendrogram). Hence, we may see under what conditions the participants' behaviour resembles one another. For example, whether or not there is any similarity among the participants when dealing with the same structure, or whether there are similarities between behaviour upon reading different structures and what sort of structures cause behaviour to vary the most.

As dependent variables, we analysed the Total Reading Time – V7 – in the region (A, B or C<sup>7</sup>) or in the context (C1, C2, C3 or C4) and the Number of Fixations – V8 – done in a region or context.

Next, we will present the various dendrograms for each text, as well as the data analysis together with a brief summary explaining the different classes that emerged. The dendrograms should be read from right to left: the sooner a class is formed, the closer (more related) are the variables in that class.

#### 3.4.1 Analysing the contexts in *Campo de Ourique*

In the dendrogram of *Campo de Ourique* **without any Deterioration at Syntactic Level (T1)**, presented in Figure 1, the formation of two large classes is clear: one of the classes contains the variables of Context 2 and Context 3; the other class gathers together the variables of Context 1 and Context 4.

If we look at these sentences' syntactic structure, we see that they are both, Context 2 and Context 3, canonical structures (Context 2 – [S V Ocl] – and Context 3 – [S V O]), where there was no constituents' movement. In contrast, in the structures in the second class (Context 1 – [que Ocl<sub>i</sub> V-t<sub>i</sub>] – and Context 4 – [WH<sub>-j</sub> V<sub>i</sub> S t<sub>i</sub>t<sub>j</sub>?]) there was constituents movement. Apart from the constituents' movement, another common feature shared by these two structures lies in the presence of a WH-morpheme.

Furthermore, in the second class two sub-classes are formed, one grouping the variables of Context 4 (with the exception of one, the Total Number of Fixations done in region A in this context (V8\_C4A), which is grouped together with the variables of Context 1), and a second sub-class grouping the variables of Context 1 (with the exception of the Total Number of Fixations are done in region A of this context (V8\_C1A), which is grouped together with the variables of Context 4). In the syntactic analysis<sup>8</sup> of these two structures we see that despite the fact that there is constituents' movement in both structures, in Context 1 the constituent that moves is the clitic complement of the Verb, while in Context 4, it is the Verb, a nuclear constituent in

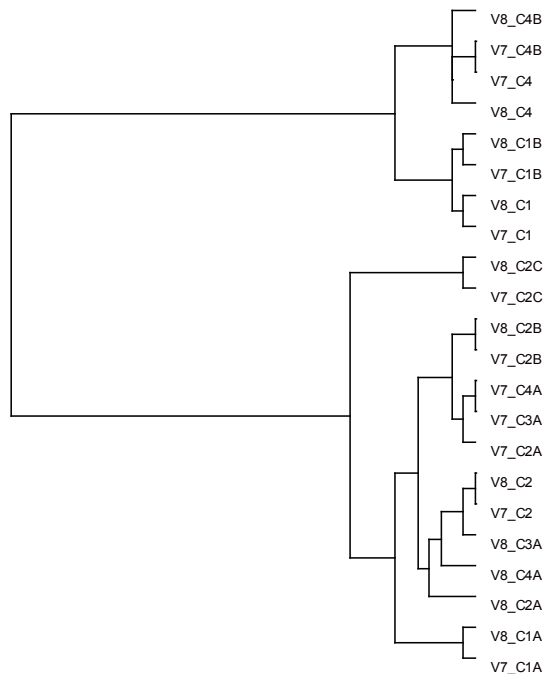
<sup>7</sup> Only Context 2 has region C.

<sup>8</sup> We do not wish to make an exhaustive syntactic analysis here but merely give a superficial description of the position the constituents occupy in the sentence.

the sentence. On the other hand, if we consider the presence of the WH-morpheme hypothesis, we see that although both structures have a WH-morpheme, in one case this morpheme is an interrogative, while in the other it is a relative. Whatever the case, at processing level, these two structures seem to be dealt with differently no matter what explanation is taken into account.

In these results, we may see that during the reading of the two canonical structures, there is no difference in eye-movement behaviours because these sentences have a very similar syntactic structure. Regarding the structures with moved constituents and with a WH-morpheme, differences distinguishing one structure from the other may be perceived in eye-movement behaviour during reading.

In a more detailed analysis, we are also able to see that the variables Total Reading Time (V7) and the variables Total Number of Fixations (V8) are nearly always grouped together with the exception of two cases. In other words, in most cases, the values obtained in the Total Reading Time in one context (or region) are linked to the Total Reading Time in another context (or region) and only in two cases do the Total Reading Time values join the Total Number of Fixations done in another region of context. These results indicate that the analysis is sensitive to the nature of the variables.



**Figure 2:** Dendrogram Campo de Ourique with Deterioration at Syntactic Level (T'1)

As in the previous dendrogram, in the dendrogram of *Campo de Ourique with Deterioration at Syntactic Level (T'1)*, in Figure 2, two large classes are evident. In the first large class are the variables of Region B (the post-problem region) of Context 4, the variables of Context 4 in general (Region A and Region B), the variables of Region B of Context 1, and the variables of Context 1 in general. Two sub-classes were also found in this class: one composed of the variables of Context 4 and the other consisting of the variables of Context 1. In the second large class there are the variables of Regions A of Contexts 1 and Context 4 and all the variables of Contexts 2 and Context 3. In these two last contexts, there was an internal movement of the constituents  $^?[V S (X) / V (X) S]$  and  $^*[S V \text{ [ ]}_{OD-cl}]^9$ , an effective movement in Context 3 and an apparent movement in Context 2. We deemed Context 2 to be an apparent movement because we believed that when the reader fails to find the internal argument after having read the verb, he/she interprets the structure as if the internal argument had been moved to a place nearer the end of the sentence. However, this interpretation becomes wrong, at the end of the sentence.

In analysing the classes that were formed, we can see that the first class gathers together the post-problem regions, where no serious changes are expected, and the variables of both Contexts, which seems to indicate that the values of the context itself are mostly a reflection of the values of Region B and not of Region A.

All the problematic regions are grouped together in the second class. If a context-by-context analysis were made, it would be seen that in Context 1 ungrammaticality occurs in Region A and it is there that it may be detected (involving local detection). The same thing happens in Context 4. Where Context 3 is concerned, the whole of the structure is ambiguous (it has only one region). Despite the ungrammaticality introduced in Region A, in Context 2, it may only be detected in the next region, that means, in Region B (we shall be talking about Region C further on), this detection has an impact both on Region A and on Region B. Hence, there seems to be a clear distinction between the region where the problem is detected and the regions where the problem is no longer felt, as in the case of Region B in Contexts 1 and Context 4.

In the dendrogram in Figure 2, contrary to what happened in Figure 1, a distinction between the different kinds of syntactic structures does not seem to exist. Nevertheless, if we examine the second class, we may see that the Region A variables in Context 1 are only grouped together with the others at a later stage. This is why there seems to be some sort of distinction between the processing of this structure and the processing of the other structures. As regards grouping Region A variables of Context 4 with the variables of Contexts 2 and 3, it seems to indicate that the structure was processed as a declarative sentence.

As for Region C in Context 2 (C2\_C), in a contrastive analysis involving the contexts in the *Campo de Ourique* text and those in *O Isolamento termo-acústico*, we

<sup>9</sup> At first, this structure is ambiguous but when reading has finished, it has become wholly ungrammatical due to the absence of a compulsory constituent. This explains the symbol  $^*$ .

verified that this structure of *Campo de Ourique* has higher values than the equivalent structure in *O Isolamento termo-acústico*. The difference is statistically significant. Therefore, the fact that it is different from all the other structures may be explained by its complexity in relation to the other structures (in this case, not syntactically but lexically and perhaps also phonologically more complicated)<sup>10</sup>.

### 3.4.2 Contrast between the *Campo de Ourique* dendrograms

By contrasting the *Campo de Ourique* dendrograms with and without DSL, we see that the dendrogram without DSL has been reorganised when compared with the text with DSL and gives rise to distinct classes being formed. However, in both dendrograms, the organisation seems to be syntactically motivated: whether or not the sentence constituents have been moved (which in the second dendrogram causes the variables in Context 1 to separate from the other variables). When analysing the first dendrogram, we considered that the grouping of Context 1 with Context 4 may be due to the presence of a WH-morpheme in both sentences. However, this hypothesis seems to be challenged by the second dendrogram as these two structures fail to get together again here (Context 1 is different from Context 4 and the remaining structures).

Anyway, as witnessed in the results, the eye-movement behaviour seems to depend upon the sentence's syntactical characteristics.

### 3.4.3 Analysing the contexts in *O isolamento termo acústico*

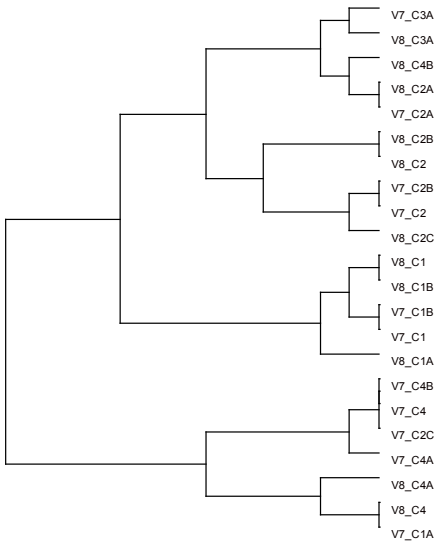
In the dendrogram of *O Isolamento termo-acústico* without DSL (T2), shown in Figure 4, three distinct classes are formed. Similar to the dendrogram of *Campo de Ourique* without DSL, the first class is made up of the variable of the two contexts containing canonical structures without the WH-morpheme, or in other words, it is composed by Context 2 ([S V O]) and Context 3 ([S V Ocl]). Similar to *Campo de Ourique* without DSL, Context 1 and Context 4 are distinct from each other, although here it emerges more clearly because each one forms an isolated class and not a subclass that belongs to the same class. It seems to confirm the hypothesis that, despite having both structures constituents' movement, the fact that the movement has to do with different constituents (nuclear *versus* non-nuclear) means that the processing of the structure is made in different ways. Another explanatory hypothesis of why the two classes have been formed lies in the presence of a WH-morpheme in both of them, an interrogative morpheme in one and a relative in the other. In general, there is a change in the behaviour of eye movements due to the internal structure of the sentence being read.

In the dendrogram of *O Isolamento termo-acústico* **with** Deterioration at Syntactic Level, in Figure 5, and similar to the others, two large classes have been

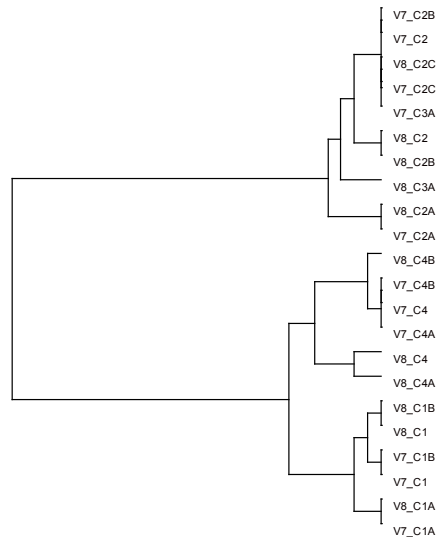
<sup>10</sup> This same structure, in Costa (1991), registered some interesting behaviour in the form of a large number of hesitations and changes in production, influenced by the word coming immediately afterwards (*casario heterogéneo*), causing [kazariu] to become [kazáriw] or [kazaráriw], for example.

formed which emerge much more clearly than in the other cases: the first is composed of the variables of Contexts 2 and 3 while the second class consists of all the variables of Contexts 1 and Context 4. In this second class, there are also two sub-classes, one consisting of all the variables of Context 1 and the other containing all the variables of Context 4.

In looking at the way the classes have been formed, we see that the variables are grouped together accordingly to the syntactic structure in which they have occurred. One of the classes is composed of structures in which the problem was not immediately detected. Another class, as in the dendrogram of *Campo de Ourique* with DSL, is made up of structures where there was no constituents' movement, which was obligatory.



**Figure 4: Dendrogram of *O Isolamento termo-acústico* without DSL.** Variables V7 (Total Reading Time) and V8 (Total Number of Fixations) in regions A, B and C in the different contexts of *O Isolamento termo-acústico* without Deterioration at Syntactic Level.



**Figure 5: Dendrogram of *O Isolamento termo-acústico* with DSL.** Variables V7 (Total Reading Time) and V8 (Total Number of Fixations) in regions A, B and C in the different contexts of *O Isolamento termo-acústico* with Deterioration at Syntactic Level.

This dendrogram shows the distinction between the classes most clearly owing to the fact that all the variables are grouped together according to the context they belong to. For example, there are no Context 1 variables in the class that mostly consists of Context 4 variables, as happens in the previous dendrograms. Only Context 2 and 3 variables are grouped in the same class, as always happened in the previous dendrograms.

### 3.4.4 Contrast between the *O Isolamento termo-acústico* dendrograms

In contrasting the dendrograms of *O Isolamento termo-acústico* with and without DSL, we see that there are similarities between the two: reading behaviours are distinguishable according to the characteristics of the syntactic structure under study, and on the basis of whether the constituents have been moved or not and what type of constituents have been moved (nuclear *versus* non-nuclear), as well as whether or not ungrammaticality / ambiguity is detected immediately. In these dendrograms, it seems likely that the WH- morpheme is once again a distinctive element, due to the fact that structures containing the WH- morpheme gather together and stand out from the other structures.

Nevertheless, despite the two dendrograms being fairly similar, they differ in matters of class cohesion. In the dendrogram showing *O Isolamento termo-acústico* with DSL, the classes become very cohesive quite early on and are composed only by the variables occurring in a single context (with the exception, as we have already pointed out, of the variables of Contexts 2 and Context 3 that are always grouped together).

### 3.5 Discussion

From the cluster analysis, we can draw some fairly interesting conclusions which may only be obtained by this sort of analysis. We may therefore conclude that the eye-movement behaviours of the participants during reading depend also upon the syntactic characteristics of the structure we are analysing, no matter what the possible explanations are: the one of movement or the one of presence/absence of the WH-morpheme. Where movement is concerned, we believe that the behaviours differ if reading has to deal with a structure with or without internal constituents' movement, the characteristics of the constituent which has been moved, and whether detection is more or less local. If we consider the second hypothesis, the one of WH-morpheme, it is assumed that eye-movement behaviour would vary if the WH-morpheme is present or not, and if the constituent is an interrogative or a relative. Whatever the case, the results allow us to confirm what we had suspected at the beginning of our analysis: eye-movements behaviour is sensitive to different syntactic structures.

In terms of the effects of altering the stimulus, the effect of changes between the two texts are more obvious in *O Isolamento termo-acústico* than in *Campo de Ourique*. This means that the TOPIC seems to be relevant in the formation of classes: the less accessible the text is, the more difficult it is to read, the greater the distinction would be between the syntactic structures.

The reorganization of the dendrograms from one text to another confirms the influence of both the DSL effect and the TOPIC effect. The DSL effect was confirmed by the reorganization from dendrograms of texts without DSL with its counterpart with DSL. The effect the TOPIC exerts, on the other hand, is observed in the contrast between *Campo de Ourique* and *O Isolamento termo-acústico*. The association of the

two variables in the more complicated text with DSL gave rise to a greater change in eye-movement behaviour.

## References

COSTA, M. A. *Leitura: Compreensão e Processamento Sintáctico*. Tese de Mestrado apresentada à Faculdade de Letras da Universidade de Lisboa, 1991.

\_\_\_\_\_. *Processamento de Frases em Português Europeu. Aspectos cognitivos e linguísticos implicados na compreensão da linguagem escrita*. Lisboa: Fundação Calouste Gulbenkian, 2005.

LABERGE, D., SAMUELS, S. J. Towards a Theory of Automatic Information Processing in Reading. In: Harry Singer and Robert Ruddell (eds.), *Theoretical Models and Processes of Reading*. 3.<sup>a</sup> Edition, Newark, Delaware: International Reading Association, 1985, pp. 689-718.

LUEGI, P. *O registo do movimento dos olhos durante a leitura de textos*. Tese de mestrado apresentada à Faculdade de Letras da Universidade de Lisboa, 2006.

\_\_\_\_\_.; COSTA, M. A., FARIA, I. H. Mover para ler: o movimento dos olhos durante a leitura de textos. *Actas do XXII Encontro Nacional da Associação Portuguesa de Linguísticas*, Lisboa, pp. 431-445, 2007.

MATIN, E. Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81(12), pp. 899-916, 1974.

SAMUELS, S. J. Towards a Theory of Automatic Information Processing in Reading: Updated. In: Harry Singer and Robert Ruddell (eds.), *Theoretical Models and Processes of Reading*. 3<sup>th</sup> Edition, Newark, Delaware: International Reading Association, 1985, pp. 719-721.

STARR, M. S., RAYNER, K. *Eye movements during reading: some current controversies*. Trends in Cognitive Sciences, 5(4), pp. 156-163. 2001.