ACCESSING REFERENTIAL INFORMATION DURING TEXT COMPOSITION : WHEN AND WHY ?

DANSAC Christophe & ALAMARGOT Denis

Laboratoire Langage et Cognition

LACO EP CNRS 1594

99 av. du Recteur Pineau

Université de Poitiers

86022 Poitiers Cedex

FRANCE

E-Mail : Denis.Alamargot@mshs.univ-poitiers.fr

Christophe.Dansac@mshs.univ-poitiers.fr

Running head : Retrieving Referential Information

Abstract

When composing a text, writers have to continually shift between content planning and content translating. This continuous shifting gives the writing activity its cyclic nature. The first section of this paper will analyse the writing process as a hierarchical cyclic activity. A methodological paradigm will be proposed for the investigation of the writing process. In the second section, we will partially present two experiments that were conducted independently, with this paradigm. Both give a coherent and interesting picture of what happens with content while the writer is planning. The characteristics of cycles depend both on the nature of the content information being recovered and on the complexity of the processes applied to this content.

A writer, when composing a text, must throughout this activity, in order to establish its content, call upon referential knowledge (knowledge about the domain, or about the particular topic). In this chapter, we will explore how content retrieval functions and suggest some theoretical and empirical answers to two questions :

The When question : is it possible to find out, during the ongoing composition, the moments when the writer needs to access this stock of knowledge ?

The Why question : what are the processes applied to content thus recovered, and are they always the same during the different moments when it is being recovered throughout the entire composition ?

1 From the retrievedcovered knowledge to the written product

1.1 The retrieval process : two processing levels

In the theoretical framework proposed by Hayes and Flower (1980) access to referential knowledge is necessary when applying two processes; during planning when this access is performed by the generating sub-process and during translating when the access to referential knowledge is conducted by the operation described as "plan next sentence : retrieve propositions" (Hayes & Flower, 1980, pp.15-16).

Considered as the first step in planning, the generating sub-process retrieves the referential knowledge from long term memory that is necessary in order to elaborate the text plan. The content thus recovered is then re-elaborated by the processes of goal setting and organising. But retrievingcovering referential content also occurs during the translating process. In this phase of composition, it is conveyed by "retrieve propositions" operation. This operation is in charge of developing (particularising) the text plan that results from planning by retrieving content that is specified in the text plan.

This distinction between two kinds of retrieval in the Hayes and Flower's model is endorsed by the distinction among the processes occurring after content is recovered. Content recovered by the generating sub-process is more likely to require re-elaboration by the organising sub-process, whereas content recovered during particularising (retrieve proposition) is more likely to be directly transformed into linguistic material by the "express next proposition part" operation.

In our view, this theoretical distinction is all the more relevant since it allows one to specify the different ways in which content is retrieved during composition. One could hypothesise that the different phases when referential content retrieval occurs throughout composition (the When) involve either one or other of these retrieval sub-processes. The kind of retrieval the writer engages in would be related to whether he is concerned with planning and / or with translating (the Why). It is an interesting empirical question whether one could detect particular points as a writer progresses through producing a piece of text that are likely to be dedicated to one or the other of these levels of retrieval.

In our opinion, a theoretical answer to these points requires an inquiry into the dynamics of composition. And more particularly here, we need to explore the dynamics of application of these two levels of retrieval. In this regard, van Dijk & Kintsch's (1983) model of written production, is relevant. Their model is similar to that of Hayes and Flower (1980; Flower and Hayes, 1980) but is more precise in its account of the dynamics of writing

1.2 Van Dijk & Kintsch's (1983) cyclic theory of writing

Van Dijk & Kintsch (1983) consider that composition activity is made up of a series of encased and iterative " strategies ". First, a " macroplanning strategy " produces a pragmatic structure that provides top-down control of the processes conducted by three other strategies named the " macrostructural strategy ", the "microstructural strategy ", and the " local coherence strategies ".

The “ macrostructural strategy ”, which is similar to the planning process in Hayes and Flower’s model,) is in charge of retrievingal and organising the main ideas of the text (note that contrary to us, van Dijk & Kintsch do not make use of the concept of idea, postulating that content is retrieved covered as semantic propositions). This function is performed by some macrostructural operations that eliminate the non-relevant ideas, and construct new semantic relations between the relevant ones. The product resulting from this processing strategy is defined as the text macrostructure.

The processes conducted by the “ microstructural strategy ”, under the control of the text macrostructure, are very similar to what is conducted by the "retrieve propositions" operation in Hayes and Flower’s model. The microstructural strategy operates the specification of the macrostructure. This is done by retrieving and ordering more specified ideas and assigning them semantic functions. Such a linearised product is defined as the text microstructure, also named text base. The text base comprise the semantic content of the text that is to be transformed into linguistic material by the fourth strategy : “ local coherence strategy ” that monitors sentence production (grammatical and lexical encoding). The function of this strategy is very similar to that proposed by Hayes & Flower through the "express next proposition part" operation within translating. Given that processing capacity is limited, van Dijk and Kintsch argue that elaborating the final written product imposes a cyclic nature on the activity. The entire writing process needs to be broken down into a cyclic execution of strategies operating upon text portions that are defined by the macrostructural content. It is worth noting that like Hayes and Flower the van Dijk and Kintsch model distinguishes two levels of content recovered, one dealing with macrostructural units and the other related to microstructural units.

Defining a processing cycle as the entire set of processes operating between two phases of content retrieval, we assume, as Van Dijk & Kintsch (1983) suggest, that cycles always start with retrieval. This is followed by macro and microstructural processing (which interest us more particularly within the framework of this chapter). Macro and microstructural processing occurs throughout the drafting of the text, by the iteration of two encased processing cycles. A macrostructural processing cycle, initiated by the activity of the generating sub-process (according to the terminology of Hayes & Flower), would be then supplemented by a succession of microstructural cycles, engaged by each application of the translating process (again in Hayes & Flower's wording). Once microstructural content is linearised and written down a new set of macro and microstructural cycles can be initiated.

1.3 Operationalising processing Cycles

Usefully, Van Dijk & Kintsch's theory permits an operational definition of the macro and microstructural products. This makes it possible, in principle at least, to identify in the written product the points at which macro and micro structural processing cycles are initiated. In order to do this, however, we need some way of identifying macrostructural features in the text. In this regard we have found the notion *semantic blocks* useful. The whole text content is constituted of several themes. We define a semantic block as the set of propositions that are developing the same theme. Thus one can consider that each semantic block comprises content related to the same macroproposition (that is, to the same main idea). So at least one macrostructural cycle will be initiated every time the writer switches to a new theme. This is because developing a new theme will necessarily require macrostructural processing of ideas relating to this new theme (the set of propositions that we are referring to as a semantic block). Microstructural cycles are executed within a theme each time the writer needs to retrieve~~cover~~ information (i.e. to particularise a macroproposition) in order to continue the text.

Thus, by applying an off-line analysis to the content of the text, it becomes possible to determine, in the textual product, where processing cycles begin. Let us note, in addition, that since the text is produced sequentially, these sites (Where) also correspond, from the on-line point of view, to different moments (When) in the writing process.

1.4 Summary

Our aim in this section was to provide a description both of the dynamics of content retrieval (the When question) and of the processes that are consecutively applied to the retrievedcovered content (the Why question). According to the theoretical suggestions of van Dijk & Kintsch (1983) one would expect the process of retrieving macrostructural content (generating) to occur before every shift in the current theme in the text in order to provide new idea units that will be dealt with by macrostructural processes (generating + organising). Conversely, developing the theme should involve the activity of a retrieval process applied to more specific content (particularising) that retrieves content likely to be processed by microstructural operations.

2 Studying retrieving and organising : some propositions

2.1 Identifying when retrieval occurs

Answering the first question (When), is not that easy, because it necessitates being able to identify operationally, within the ongoing activity, the moments when the writer needs to access referential knowledge. It is of course possible to describe a posteriori the presence of such a mental activity by interpreting the verbal protocols of a writer or by regularly questioning him on the nature of his processing in progress (retrospection method used for instance by Kellogg, 1987). Although frequently used, these methods of investigation have a number of disadvantages: interpretation of protocols is necessarily subjective or they impose the segmentation of the activity, or even they place the writer in a dual task situation.

We propose as a methodological solution to this problem to have the writer produce a text about an image, which we will call the imaged referent. If the writer needs to break off translation in order to see the image, one can detect, during his activity, the moments when he needs to access referential information (information that writers are required to base their text upon). Also let us note that the recourse to an image as the referent for the composition comprises advantages within the framework of an experimental investigation : it allows the control of the information contained in the referent (then imposed on writers and homogeneous to them) as well as the possibility of varying the nature of the referent (variations of the information contained in the image).

Such a principle can be realised using a computer device like G-Studio 2.0 software (for a detailed presentation, see Chesnet, Guillabert & Espéret, 1994). The writer is asked to compose a first polished draft, writing on a digitising tablet with a special pen. By pointing to a specific area on the tablet with the pen (see figure 1), she can have the software displaying the image on the computer screen. With such a device, she has to break off translation to see the image she is composing about. G-Studio 2.0 allows the on-line recording of every movement of the pen upon the tablet. We can then replay the pen movements and identify, amongst other things, when during the writing process the writer consulted the image.

Insert figure 1 about here

We assume that moments when the writer needs to see the image (i.e. to encode information from the referent) are in some senses equivalent to moments where the writing activity would require the retrieval of referential information from long term memory if the writer had to produce the text only from memory.

2.2 Identifying why retrieval occurs when it does

Once these consultation phases are located in the text, analysing their duration can help us to find out answers to the Why question. Much work on oral production (e.g. Goldman-Eisler, 1972 ;

Butterworth, 1980) and written production (e.g. Matsuhashi, 1981, 1982) is based upon the study of pause duration. " Silent " phases in the course of production have been interpreted as an indicator of planning processes (Kowal and O'Connell, 1987; Foulin, 1995). In the same way we will assume that the length of time that a person in one of our studies spends looking at the imaged referent is a measure of the number or complexity of the underlying retrieval processes. Shorter consultations indicate ~~processing, ensuring~~ at least th~~a~~te ~~new~~ information is encoded and th~~e~~ ~~at th~~i~~ts~~ is transformed into a linguistic product (microstructural processing and sentence generation). Longer consultations suggest the execution of additional, macrostructural processes that involve reorganising the retrieved content before its translation.

We are arguing, therefore, that processing cycles are hierarchically organised, that they are associated with text structure, and that the period of time spent consulting the imaged referent is proportional to the number and complexity of operations conducted. If these assumptions are correct then we should observe a relationship between consultation times associated with the application of content retrieval in relation to the level of representation in the text at the point at which the cycle is initiated.

Processing phases that precede a semantic block, since they involve organising, will last longer than processing phases where the content is recovered within the blocks, because the latter involve only translating. Shorter processing phases in-blocks are due both to there being no need to organise, and to the fact that translating processes are largely automatic at least in adults (cf. Levelt, 1989 ; Bock, 1982 ; see Mc Cutchen, et al. 1994 for an analysis of this point). The prewriting phase before the writer starts producing text on the page should be the longest of all as this involves not only macro-planning ~~-~~but also general exploration and retrieval of the information to be reported.

If such a temporal pattern can be found through an on-line analysis of the composition, then this will provide support for the van Dijk & Kintsch's (1983) theory. But in order to find more evidence of the fact that the macrostructural processes operate primarily before semantic blocks, we propose two methods of experimental investigation that were used in the experiments reported here : 1) to vary, upstream retrieval~~covery~~, the prior organisation of referential information within the image ; and 2) downstream of retrieval~~covery~~, to analyse the more or less complex activity of the organising process which operated on the same referential information.

2.2.1 Varying prior organisation of referential information.

Previous researches have already shown that macrostructural processes, either generating or organising, are facilitated when referential knowledge is more available or better structured within long term memory. For instance, Kellogg (1987) demonstrated (by measuring means of secondary task reaction times) that the retrieval and organising processes are required less of a writer who is an expert in the referential domain. This facilitation from referential expertise has also been found by Caccamise (1987). Examining the quality of the products, she found that experts produced more ideas with greater hierarchical organisation of their ideas (see also Voss, Vesonder & Spilich, 1980 ; Benton et al. , 1995). This facilitation can be explained by the fact that experts' knowledge bases are structured such that there are a larger number of semantic relations which serve to chunk knowledge units (Egan & Schwartz, 1979 ; Dempster, 1981 ; Alamargot, 1997). Such chunks (as defined classically by Miller, 1956) facilitate retrieval as well as the encoding processes (through easier storing in working memory)~~.~~ Since chunks structure referential knowledge at hand during production, they also facilitate organisational processing during composition as Kellogg (1987) and Caccamise (1987) demonstrated: ~~.~~ If knowledge is well organised it does not need further elaboration.

Consequently, within the framework of our paradigm, experimentally varying the prior organisation of referential information comprised in the image may be an interesting way of assessing the necessity of macrostructural processing after content retrieval. In other words, one way to test the actual execution of the organising process, before a semantic block (i.e. before a new theme) begins to be written, is to vary prior organisation of referential information. Then the time required for processing referential information before semantic blocks (the macrostructural process of generating and organising) should be influenced by previous organisation, whereas processing phases within blocks should be exempt from such an influence, since microstructural processes do not involve organising operations.

2.2.2 Contrasting products obtained from the same referential information.

Our second approach to testing the hypothesis that organising occurs during the macrostructural phase of content retrieval is based upon another principle. The idea is to fix the characteristics of referential content among all writers and to analyse, by means of a post hoc content analysis of each text, which organising operations were conducted upon this content. Such an analysis can lead to a quantitative indicator of the amount of operations achieved. By differentiating between organising operations that are more or less complex, each “ kind ” of product can be related to the time that the writer spent on its elaboration.

If the hypothetical model of the activity described above is applied some expectations can be formulated. The phase where retrieval applies to macrostructural content (i.e. before a semantic block) should have a duration in relation to the amount and complexity of the macrostructural operations applied to the content. On the contrary, this relation should not be found with the phase where retrieval consists in particularising (i.e. within a block).

The two experiments described in the next section of this chapter were conducted independently, addressing the when and why questions with either one or other of the above-proposed methods of investigation (for more details, see respectively Dansac & Passerault, 1996; Alamargot, 1997). Both were designed to explore the hierarchical and cyclic aspect of the writing process. In experiment 1, we used the first method described above. We manipulated the organisation of referential information ; this was done via the location of economic features on a fictitious map. The second experiment used the second method, aiming at showing how processing times can be explained by the nature of differing organising operations conducted upon the same imaged referent.

3 Experiment 1: varying prior organisation of referential information

3.1 Method

We provided subjects with a map of an imaginary island (see figure 2). This map showed economic resources that belonged in equal number to three categories (agriculture, industry and transportation) with the same number of resources in each of the three regions.

Insert Figure 2 about here

The resources on the map were either "ill-grouped" or "well-grouped" (between-subject factor). In the ill-grouped version, the categories of resources were randomly assigned to the three regions (see table 1). In the well-grouped version, they were so arranged that every category of resources was mentioned only in two regions. As stated in the first section, we assume that the generating sub-process is facilitated in this second case, i.e. when information is easier to chunk, because resources may be grouped easily within each region according to their category.

The subjects had to write a text about the fictitious map. In order to make the writing task more complex as regards the organising processes, the writing assignment mentioned that the goal of the text was to compare the regions with regard to their economic characteristics.

Insert Table 1 about here

One hundred and twenty three psychology students (19 to 25 years old) participated in this experiment. All were volunteers. Subjects were provided with the writing assignment and had to produce a single, polished draft. They were asked to write their text on a digitising tablet, controlled by the G-Studio 2.0 software described above. So as to see the map on the computer screen subjects had to break off translation and point to a special area of the tablet with the pen (see figure 1). They could consult the map as often as they needed, and had no time limitation to complete their composition. They were required to produce a page of text.

3.2 Analysis

The length of the texts was computed in both number of sentences and number of words. The two groups wrote texts of similar length : 199.8 (s.d. = 48.2) words for the ill-grouped version and 201.0 (s.d. = 52.3) words for the well-grouped version

We divided each text into semantic blocks, each block beginning when the writer introduced new thematic information. In this task, therefore, a semantic block is found each time the text introduces either a new region or a new economic category . Block boundaries, with a few exceptions, coincided with sentence boundaries. We analysed the duration of map consultations coding them as occurring in one of three locations: between blocks, between sentences within blocks, or within sentences within blocks.

3.3 Results

In order to illustrate how the paradigm functions, we present in table 2 an analysis of the data for a representative subject from the ill-grouped condition which has been translated word for word from French.

Insert Table 2 about here

As can be seen in table 2, block boundaries (signalled by the $ marks) are very often the place for a long consultation of the map (consultations signalled by ⊗ marks #1, 2, 6, 9). These consultation phases are presumed to reflect macrostructural planning. Within the blocks, we can observe an interesting pattern of information retrieval. The writer expresses the macroproposition she constructed (e.g., " These three regions also differ regarding their agricultural productions.") and then returns to the map in order to find its actual content (consultation #7). Four times (#4, 5, 8, 10) this subject begins to formulate the content goal of the statement (e.g. " and in region C a "…) and then calls for the map in order to recover this content.

Thus the pattern of access to referential information is highly compatible with the assumptions of a cyclic model involving macropropositions and their particularising. By dividing up the writing process in this way the writer avoided loading her working memory with content that was not going to be formulated immediately.

The mean duration of consultations (see table 3) was a function of their location in the text, and this duration varied according to the nature of the unit that followed the consultation.

Insert Table 3 about here

The MANOVA performed on mean consultation times showed a strong effect of the nature of the map. For all locations, the mean duration was lower for the well-grouped map than for the ill-grouped one ($F(4,118)=3.04$, $p=.020$). Univariate analyses showed a more contrasted picture. Within blocks, either within or between sentences, consultations were significantly shorter in the well grouped condition [$F(1, 121)=4,79$, $p<.05$ for between sentences and $F(1, 121)=6,74$, $p<.025$) for within sentences]. However this effect was only marginally significant for consultation prior to any transcription ($F(1,121)=3.43$, $p=0.06$, and for pre-block consultation duration, the difference between ill-grouped and well-grouped was not significant($F(1, 121)=1,36$, $p>.05$).

3.4 Discussion

The findings from experiment 1 fit a hierarchical cyclic model relatively well. The model proposed by Kintsch & Van Dijk (1983) is accurate in predicting the duration of an access to referential information as a function of its location within the semantic structure of the text.

In our results, pre-writing duration appears to be the longest. This duration may be influenced to some extent by chunking facility, although the effect of information organisation was not very great and only bordered on significance. The particular status of this initial pause makes it difficult to interpret its duration. It also reveals assignment comprehension, task environment analysis, establishment of pragmatic goals of the text and, moreover, the first exploration of the referent to be described that is likely to be the source of the effect we observed.

More interestingly, pre-block consultations were about 50% longer than consultations within blocks. These results replicate those of Savigny (1995) and Alamargot (1997, see also experiment 2). Such a superiority suggest that pre-block consultations reflect the execution of more numerous or complex processes than in-block consultations do. In-block consultations are likely to be traces of retrieval phases that are associated only with translating. In other words, macrostructural processes that operate before a block is translated are typically initiated by a retrieval phase. This phase is then followed by a long phase during which this newly retrieved macrostructural content is organised. During translation of the block the retrieval process would be more concerned with local content (particularising). This new content is likely to be transcribed automatically.

If this interpretation is correct, one would expect organisation of the information to be written about, which facilitates chunking, to play its major role on the pre-block rather than within-block consultations. When pieces of content can easily be chunked, the cognitive demands of their organisational processing are reduced. We find here that the organisation of the information

presented to the writers had an effect on the duration of consultations at every location but that this effect is only reliable for consultations that occurred within blocks. The temporal difference observed between the groups can come neither from the text content nor from its organisation, since they are globally similar among the texts produced**.** So we have to admit that writers, confronted with the ill-grouped version of the map (without chunk facilitation), needed more consultation time to obtain similar products.

What is interesting is that this increase occurs within the semantic blocks. We hypothesised that in-block consultations were just devoted to retrievecover information required by the translating process ; and that consequently, these consultations were unlikely to reveal content organisation, contrarily to pre-block consultations. Taking these results into account, we must then reject both assumptions. We must consider at least that some time-consuming processes occur in addition to formulation (within the semantic blocks) when information is ill-grouped. We assume that well-grouped information facilitates chunking. Chunking activity involves the possibility of retrieving, a larger set of information within a single processing cycle. Conversely, because processing capacity is limited, if chunk elaboration is not facilitated, as was the case in the ill-grouped condition, the whole information set cannot be recovered within a single processing cycle. In this situation, information retrieval and organising are also necessary during within-block processing cycles which would normally be devoted purely to translation. That is, whenever content to be expressed cannot be chunked in working memory, writers need to distribute retrieval and organising within a macrostructural cycle. This distribution leads them to execute longer and more complex processes applying to the information retrievecovered during translating.

Our interpretation of the phenomenon observed here is based on the assumption that chunking plays an important role in facilitating information retrieval and short term storage. When information can be readily chunked, all of the content necessary for a macrostructural cycle can be

retriecoveredbefore the transcription of a new semantic block begins. This suggest not that writers will never need to retrievecover information during translation, but that when within block retrieval is necessary it will occur relatively effortlessly because of structured nature of the information to be retrieved.

4 Experiment 2: contrasting differing organising products

Different content therefore leads to different organisation of the writing process. One can also wonder whether, given a same set of information, writers who produce differing content organisation also present differing partitions of processes. As varying retrieval costs can lead to delayed processes, one could also expect the organising process to modify the characteristics of processing cycles according to the complexity of the operations it carries out. Having seen the indications given by the results from experiment one, we can hypothesise that when the organising operations applying to content units are more complex to perform, the processes involved are accordingly more delayed within semantic blocks. Experiment two tried to support this assumption.

While organising content (establishing the macrostructure of the text), the writer must carry out several operations upon the semantic units that are retrieveoveredded. According to Hayes & Flower (1980), these operations include evaluating (judging appropriateness) ordering (choosing sequence…) and relating content (finding connections, identifying categories…). We will focus here on the last operation, relating content and particularly processes involved in making comparisons. Two specific types of comparisons may be distinguished (Alamargot, 1997) : “ lLinked cComparisons ” and “ iIntegrated cComparisons ”.

Linked comparison*s* correspond to the operation that Hayes and Flower describe as “ subordinating a new topic to a previous one ”. In our view, this involves developing a new theme by referring to information contained in a theme that is already present in the text that has been

produced so far. That is, when establishing a linked comparison the writer relates the content pertaining to the new semantic block to the content already translated in a previous block (e.g., “the second region contains a port like the first one, but this time it is devoted to importation and exportation ”).

Integrated comparisons build another kind of relation between topics. This comparison is achieved via the identification of a conceptual category which permits the grouping of different topics (the operation "identify a category" in Hayes & Flower, 1980). The topics, in this case can be seen as elements of this category. Integrated comparison implies constructing the content of a semantic block through simultaneous processing of two or more new topics (e.g. “ Northeast and Northwest benefit from harbours but the latter is concerned with merchandise whereas the former is devoted to pleasure sailing ”).

Before actually making a comparison the writer searches for features that the two topics have in common. This will be true whichever type of comparison is to be made. Once these features have been identified, however, the actual processes adopted will depend upon the nature of the comparison to be made. It is probable that the processes associated with making linked comparisons are less resource demanding than the processes associated with making integrated comparisons. This difference is due to the number of features to be rehearsed. In the case of integrated comparisons, all the features (even non relevant ones) must be rehearsed because each one is involved in the comparison.. However, when a linked comparison is executed the writer needs to keep activated only the features which differ between topics. Hence, as we are able to estimate the number of such relations by examining the content of a text, we might succeed in finding out which processing phase handles the elaboration of these relations. Moreover, assuming that making an integrated comparison is more complex than

making a linked comparison we should expect integrated comparisons to be operated whilst translating a semantic block and linked comparisons to be operated before the block is translated.

The aim of experiment 2 was twofold. First, it was conducted to verify the relationship between the length of content processing phases and the point at which they occur within the text that we identified in the first experiment. Second and critically for our present concern, we wanted to explore the relationship between processing time and the ways in which writers make associations between ideas in the text.

4.1 Method

Twenty four psychology students participated in this second experiment. They were aged from 20 to 35 ; all were volunteers.

The paradigm was essentially the same as in experiment 1. The referent was a set of five geometrical figures (see figure 3) that subjects were asked to describe.

Insert Figure 3 about here

As in experiment one, they could see the set of figures on a computer screen as often as they needed, and their writing activity was recorded on a digitising tablet. The writing assignment was " to explain to a reader the structure of these figures in such a way that he would be able to draw them ". The figures used were all composed of the same squares, triangles and trapeziums. They were conceived in a manner that made it possible to describe them by comparing their features (finding relations between each one's height and shape...).

4.2 Analysis

The text produced was segmented into semantic blocks. The criteria were the same as previously; a new block was identified every time a new figure was introduced as the current theme. The text was also analysed for evidence of linked and integrated comparisons.

A linked comparisons was identified when a writer described a figure and quoted the features it shared with another that had already been described. (e.g., " this figure is the same as the previous one, except that the locations of the small squares and the small triangles are inverted "). An integrated comparisons was identified when a writer simultaneously described several figures by identifying the characteristics that they shared (e.g., " the right-hand figure and the left-hand one are both made up of a trapezium on which a small square and a small triangle are disposed "). All the texts collected were double coded by independent judges. Inter-judges reliability was high both for number of linked comparisons ($r=.979$ $p<.001$), and for number of integrated comparisons ($r=.983$ $p<.001$).

We also analysed the duration of figure consultations in relation to their location within the text : before text, before a semantic block and within a semantic block (in this second study no distinction was made between the referential accesses occurring at sentence boundaries and the ones occurring within sentences). The number of linked comparisons and of integrated comparisons produced in the texts were noted.

4.3 Results

The mean consultation times (see figure 4) confirmed the results of the first experiment : the hierarchy of their duration as a function of location in the semantic structure of the product was replicated.

Insert Figure 4 about here

The main effect of the kind of consultation was significant (Friedman ANOVA X^2 (24/2) = 28.58 ; $p<.001$). More precisely, pre-text consultation time (mean=56 sec) was greater than pre-block duration (mean=30.3 sec : Wilcoxon $z = 3.22$; $p<.001$) that was greater than in-block duration (mean=19,3 sec : Wilcoxon $z = 3.54$; $p<.001$).

These results validate the pattern encountered in experiment 1. The processing time during consultations occurring at the start of semantic blocks is greater than those occurring within blocks. When the writer is asked to communicate a very different kind of information, the experimental paradigm therefore gives the same coherent result. Once again the pattern of consultation times is compatible with the hierarchical cyclic model of Van Dijk & Kintsch (1983). Pre-block consultations, we would argue, are associated with macrostructural planning while in-blocks ones are associated with microstructural processing. As expected, macrostructural planning is more complex and associated with longer consultations than the microstructural processing.

We conducted two multiple linear regression analyses to explore the relationship between the number of comparisons, either linked or integrated (as the dependant variables), and processing time during consultations occurring at the different locations (as independent~~independant~~ variables). It depicts very interestingly what occurs during these consultations. First it should be noted that pre-text duration does not explain any significant part of the variance, whatever the kind of relation. Concerning the first regression analysis, the variance of the amount of linked comparisons inserted was only ~~be~~ (marginally) explained by the variation of pre-block consultation times (13% of variance explained, semi-partial r = .37 ; p=.07). On the contrary, according to the regression operated upon the number of integrated comparisons, 24% of the variance of the amount of such comparisons is explained by the variation in in-block consultation times (semi-partial r = .49 ; p=.01). Although 11.6% of this variance would be attributed to the variations of pre-block consultation times (semi-partial r = -.34 ; p=.08), this is with a negative influence of increasing consultation times upon the insertion of such more elaborated comparisons.

Thus, an increasing number of linked comparisons should be attributed to an augmentation of the consultation time occurring before a semantic block. That is, linked comparisons depend on the duration of the macrostructural planning phase. But when a subject manages to insert integrated

comparisons in his/her text, the amount of such comparisons is clearly directly linked to an increasing processing time during in-block consultations. Moreover this contribution is at the expense of the processing time before a block. In other words, when making more integrated comparisons the writer reduces the amount of time he or she spends processing information before a semantic block and augment the processing time within blocks.

4.4 Discussion

These results therefore support the idea that organising processes occur at different times according to the nature of the relations the writer builds upon content. When content pertaining to a semantic block is to be compared with content already translated in a previous block (a linked comparison), this relationship is established before the beginning of the block translation. Hence this kind of connection will occur during the retrieval phase preceding a semantic block. Consequently, making linked comparisons will increase the duration of pre-block consultations. In the case of integrated comparisons, before starting to translate a semantic block, the writer only needs to identify a category that encompasses both of the concepts to be expressed. This identification will serve as a basis for the construction of a macroproposition (e.g. “figures are similar because...”). It is only during the translation of this macroproposition, while its content is being particularised, that the writer will actually have to relate the features of the two exemplars of the category (“ the shapes of the two figures are the same size ”). This will lead to an increase in time spent in microstructural planning. This raises a new issue: why does establishing this kind of relation mainly depend upon retrieval that occurs within semantic blocks?

According to the cyclic model of Van Dijk & Kintsch (1983), there is no reason to think that a whole integrated comparison cannot be planned before a semantic block, as linked comparisons are. We suppose that if a writer who is realising integrated comparisons needs to delay

corresponding operations, it is because identifying a category implies holding all the shared and differing features of the two figures implicated in the comparison. Due to working memory limitation, it may be that holding and processing all these features cannot be managed entirely during the macrostructural consultation preceding a block. As in experiment 1, when there was no chunk, it seems that limited processing capacity necessitates the postponement of some costly operations. Linked comparisons, which require fewerseveral attributes to be held simultaneously, are made as soon as the macrostructural cycle is initiated.

Thus it appears that the complexity of the operations associated with organising may also influence the characteristics of processing cycles. Depending upon the way in which the content to be expressed is manipulated, these operations may be realised either during the planning of the entire semantic block or during the translating of the block. Our results therefore suggest that writers balance planning content before versus within a semantic block.

5 General Discussion

When does a writer access Referential Information ? We assume that access to the visually-represented information in the experiments that we have described is analogous to retrieval from long term memory when the knowledge to be expressed is in the writers head. If this is the case then our findings suggest that content retrieval occurs repeatedly throughout the writing process. Content retrieval may occur at the beginning of a macrostructural cycle as well as at the beginning of a microstructural cycle. The single case analysis described in the first experiment does particularly suggest that one could predict, as the text content develops, when a writer will need to retrievecover referential information. Being able to predict these moments could be an interesting issue in specifying the answers to the *when* question.

Of greater interest are the *why* answers suggested here. In introducing this chapter we suggested two levels at which content is retrieved: during the generating sub-process that contribute

to macrostructural processing, and during particularising that is associated with microstructural processing. Both macrostructural and microstructural planning were conceived as occurring within encased cycles that were initiated respectively by generating and particularising. The on-line course of composition and the duration of consultations observed here support this model. However, the results of our experiments also suggest that the function of processing cycles may vary. What is usually expected as a translating phase may become, under particular circumstances, a cycle which focuses on organising. One could interpret the modification of the processes partition throughout the cycles as ~~seeming to~~ being related to the cost of retrieval and of organising. It seems that, as a general rule, planning content is executed before the translation of a new theme only when the processes involved are effortless. Thus, our research suggest that the distribution of content retrieval during the writing process is dependent, in part at least on (a) the prior organisation of referential knowledge which affects the ease with which content can be chunked, and (b) the complexity of the relationships that the writer builds and represents in the text.

When content retrieval is facilitated by prior organisation that chunks concepts or when organising operates only on a small number of semantic units as is the case when linked comparisons are made, content retrieval may occur only at the start of a macrostructural cycle. Thus the whole thematic content may be planned before any translation occurs. However, if referential knowledge cannot be chunked, or when more complex organising occurs, some content planning cannot be conducted before the theme begins to be written down. In this case, encased cycles usually devoted to translating appears to be weighted with additional processes.

Whether increased processing costs be incurred by retrieval or by organising, it should be noted that this increase is related to a greater amount of information held within working memory. Therefore, since the working memory span fixes the number of units that can be kept for computation, it is very likely to be the prevalent factor in determining the partition of processes

among cycles. Indeed, according to Bereiter & Scardamalia (1987), composing a highly coherent and organised text is difficult in that it requires the cognitive system to activate simultaneously, in working memory, more information in order to organise the whole content. This working memory accounting for more complex writing strategies is provided with support here : even in adults, memory span places such strong constraints on the activity that writers must adapt their processing partition in order to achieve content planning. Too many demanding organisational processes provoke cognitive overload. This overload requires the writer to postpone some organising within a translation phase that is thus slowed.

This interpretation based on the cost of organisational processes is tempting, nevertheless the paradigm used here does not allow us to validate it definitely. This paradigm elucidates some classically non-interpretable pauses : in this situation they correspond to a need for referential information. Nevertheless, the paradigm as such does not give any indication of the demand the processes place on the cognitive resources. Thus, increasing consultation time testifies to the execution of more complex or more numerous processes ; it does not warrant that these processes be~~be~~ more resource-demanding. To endorse this~~our~~ interpretation, secondary tasks should be associated to our mode of pause analysis. Then altered performance in the secondary task during consultations would be evidence that the processes are not only more complex or numerous but also more resource-demanding.

The research presented here is somewhat exploratory. Both experiments aimed at testing this on-line paradigm as a means of investigating the *When* and *Why* questions. It appears that this paradigm gives interesting insights into the retrieval process in relation to the processes consecutively applied. The results are encouraging and lead to the consideration that this paradigm could be fruitful in the study of the dynamics of the writing process. One of its limits, however, may consist of its lack of "ecology" : the subjects must write without preliminary draft, the referent

is imposed, initially unknown, and its nature involves the composition of a descriptive text. We believe however that the phenomena observed in this situation can be generalised with other types of texts. Indeed, there is no reason to suppose that the nature of the writing processes differs from one situation to the other, even if their resource demand and their relative importance may differ. A possible extension of this paradigm will be to test various types of referents, or type of text, in order to find out the other parameters that actually influence the dynamics of composition.

References

Alamargot, D. (1997). *Processus de récupération et d'organisation dans l'activité de rédaction de texte : effets de l'acquisition de connaissances référentielles*. PhD Dissertation. Université of Poitiers, France, January.

Benton, S.L., Corkill, A.J., Sharp, J.M., Downey , R.G. & Khramtsova, I. (1995). Knowledge, interest and narrative writing. *Journal of Educational Psychology, 87*, (1), 66-79.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition* . Hillsdale, N.J. : L.E.A.

Bock, J.K. (1982). Toward a cognitive psychology of syntax information processing contributions to sentence formulation. *Psychological Review, 89* (1), 1-47.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production* (pp. 155-176). New York : Academic Press.

Caccamise, D. J. (1987). Idea generation in writing. In A. Matsuhashi (Ed.), *Writing in real time : modelling production processes* (pp. 224-253). Norwood, N.J. : Ablex.

Chesnet, D., Guillabert, F. & Espéret, E. (1994). G-Studio : un logiciel pour l'étude en temps réel des paramètres temporels de la production écrite. *Année Psychologique, 94* , 115-125.

Dansac, C. & Passerault, J.M. (1996, August). *Récupération, organisation hiérarchique et séquentielle dans le décours temporel de la production*. Oral communication at the Congrès International de Psychologie. Montréal, August 16-23.

Dempster, F.N. (1981). Memory span : sources of individual and develomental differences. *Psychological Bulletin, 89* (1), 63-100.

Egan, D.E. & Schwartz, B.J. (1979). Chunking in recall of symbolics drawings. *Memory and Cognition, 7* (2), 149-158.

Foulin, J.-N. (1995). Pauses et débits: les indicateurs temporels de la production écrite. *L'Année Psychologique, 95* (3), 483-504.

Flower, L. S. & Hayes, J. R. (1980). The dynamic of composing : making plans and juggling constraints. In L. W. Gregg, & E. R. Steinberg (Eds.). *Cognitive processes in writing* (pp. 31-50). Hillsdale, N.J. : L.E.A.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and speech, 15* (2), 103-113.

Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, N.J. : L.E.A.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition, 15* (3), 256-266.

Kowal, S., & O'Connell, D. C. (1987). Writing as language behavior: Myths, models, methods. In A. Matsuhashi (Ed.), *Writing in real time* (pp. 108-132). Norwood, New Jersey: Ablex.

Levelt, W. J. M. (1989). *Speaking : from intention to articulation* . Cambridge, Mass. : The M.I.T. Press.

Matsuhashi, A. (1981). Pausing and planning : the tempo of written discourse production. *Research in the teaching of English, 15* (2), 113-134.

Matsuhashi, A. (1982). Exploration in real-time production of written discourse. In M. Nystrand (Ed), *What writers know : the language and structure of written discourse* (pp.269-289). New York : Academic.

Mc Cutchen, D, Covill, A., Hoyne, S.H. & Mildes, K. (1994). Individual differences in writing : implication of translating fluency. *Journal of Educational Psychology, 86*, (2), 256-266.

Miller, G.A. (1956). The magical number seven, plus or minor two : some limits in our capacity for processing information. *Psychological Review, 63*, 81-97.

Savigny, A. (1995). *L'organisation des connaissances et la production de textes procéduraux (études off-line et on-line)*. PhD Dissertation. Université of Poitiers, France, Avril.

van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension* . New-York : Academic Press.

Voss, J.F., Vesonder, G.T. & Spilich, G.T. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning and Verbal Behavior, 17*, 651-667.

Author Notes

The two authors contributed equally to this paper.

We would like to thank Mark Torrance, David Galbraith and anonymous reviewers for very constructive comments on an earlier version of this chapter.

Table 1

Information structure on the map

		Region		
		North-west	North-east	South
ill-groupedG-	agriculture	bovines	ovine	vineyards cereals
	transportation	leisure navigation airfield	freight navigation	airport
	industry	textile	chemistry	automobile electricity
well-groupedG+	agriculture	ovine bovines		vineyards cereals
	transportation	leisure navigation airfield	freight navigation airport	
	industry		textile chemistry	automobile electricity

Table 2

An example of protocol and its analysis

The text produced The physical characteristics of the written product were respected within the limits given by the use of a word processor. $ and ⊗ marks were added to signal respectively our segmentation and the map consultations. The translation from french is made word for word as exactly as possible.	**Our notation** Consultation duration (sec.) and nature (PT : pre-text ; PB : pre-block ; PS : in-block pre-sentence ; IS : in- block in-sentence)
$0 ⊗0This island is constituted of three regions which are differentiated both on their location and on the resources they produce. Each region groups two poles where are developed the different resources	⊗0 *: 98.0 s (PT)* **Introduction**
$1⊗1Contrary to the region B, the regions A and C contain a harbour : In the region C : it is most a harbour devoted to importing and exporting merchandises while in the region A is located a harbour rather for pleasure which receive sailing boats.	⊗1 *: 30.0 (PB)* **Harbours block**
$2⊗2These three regions A, B, C have industrial activities which are different.	⊗2 *: 8.8 (PB)* **Industries Block**
⊗3 In the region A one finds textile industry, in the region B ⊗4 industry of automobile and electricity, and in the region C a⊗5 chemistry industry.	⊗3 *: 8.7 (PS)* ⊗4 *: 5.4 (IS)* ⊗5 *: 1.2 (IS)*
$3⊗6These three regions also differ regarding their agricultural productions.	⊗6 *: 18.2 (PB)* **Farming Block**
⊗7 Indeed, the region B cultivate wheat whereas the region C cultivate vineyards for the production of wine.	⊗7 *: 13.6 (PS)*
$ 4 The breeding of animals are of different nature in the region A ⊗8 one mostly breeds cows while in the region C, one breeds sheep.	**Breeding Block** ⊗8 *: 0.6 (IS)*
$5⊗9This island contain at last two airports. In the region A ⊗10, the airport is devoted to travellers while in the region B the airport is used as an area of import and exportation of merchandises.	⊗9 *: 14.8 (PB)* **Airports Block** ⊗10 *: 3.5 (IS)*
	- 208 words, **- 11 sentences** **- 4 comparisons** **- economic-oriented structure**

Table 3

Mean across subjects of mean consultation time (in seconds) at each kind of location

organisation of referential information	Location			
	pre text	pre block	within block, pre-sentence	within block, within sentence
ill grouped	87.8	19.4	15.1	12.1
well grouped	74.3	16.3	8.9	7.4
overall (s.d.)	80.8 (40.7)	17.8 (15.1)	11.9 (15.7)	9.7 (10.1)

Figure 1

Digitising tablet and iImaged rReferent display

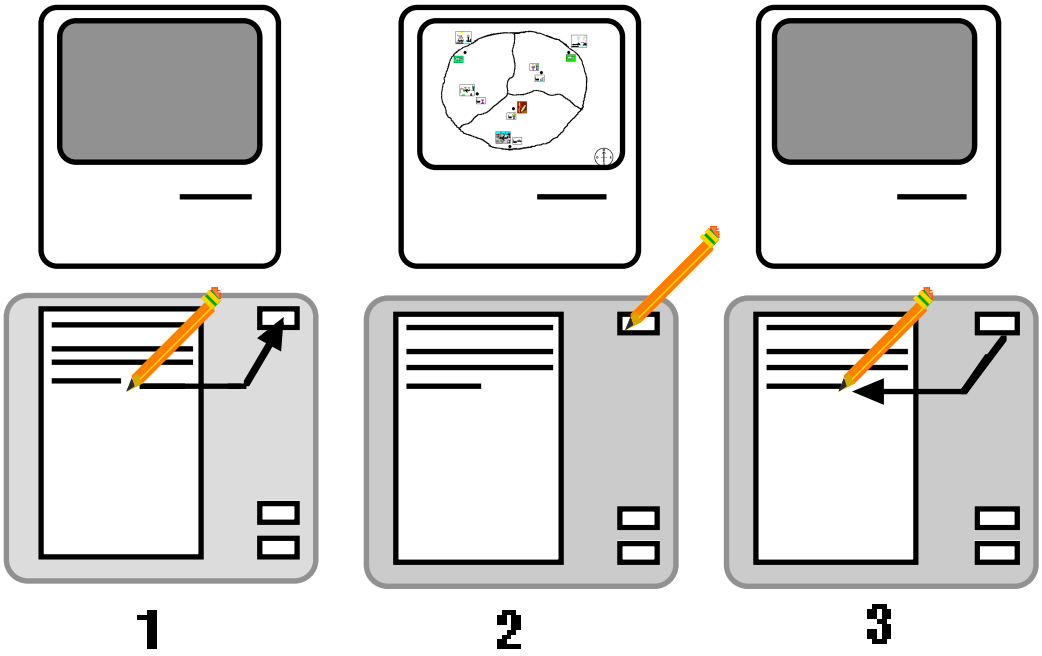

Figure 2

The island to be described : the ill-grouped version

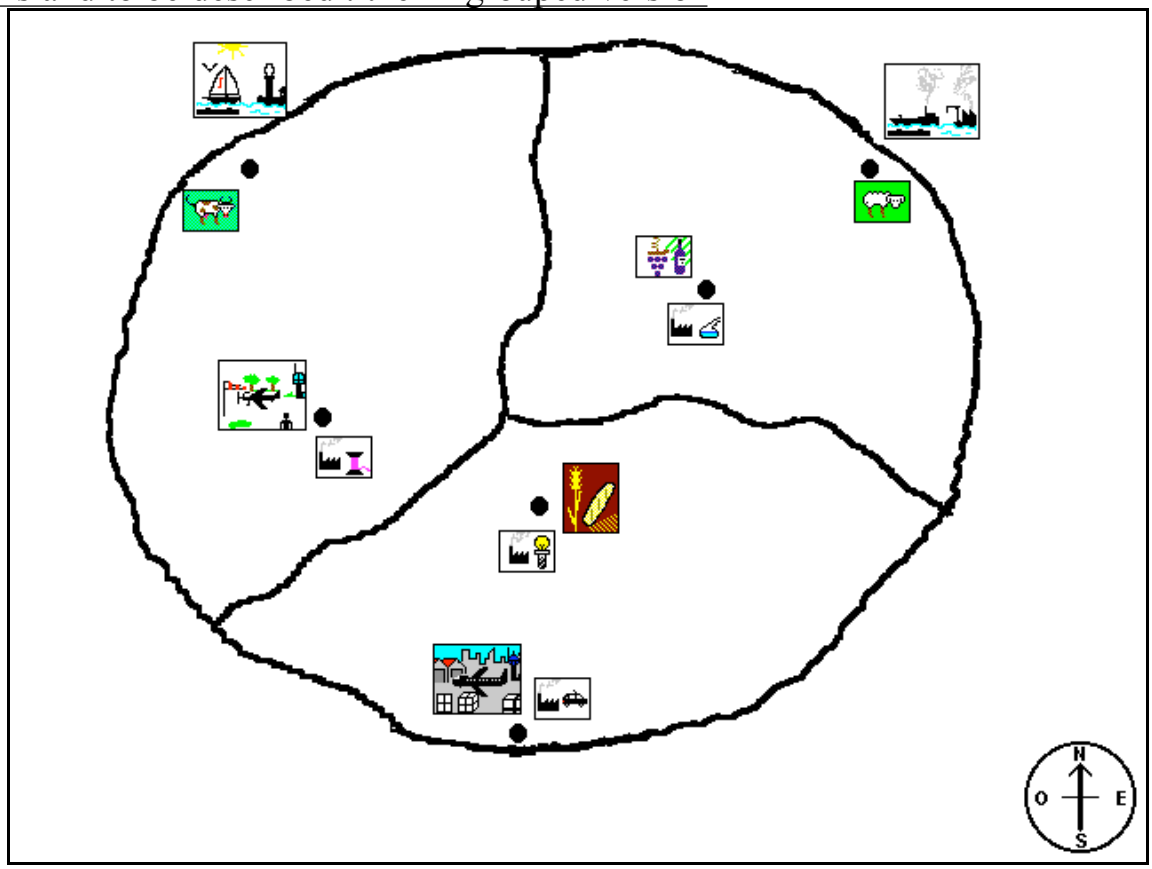

Figure 3

The iImaged rReferent in experiment 2

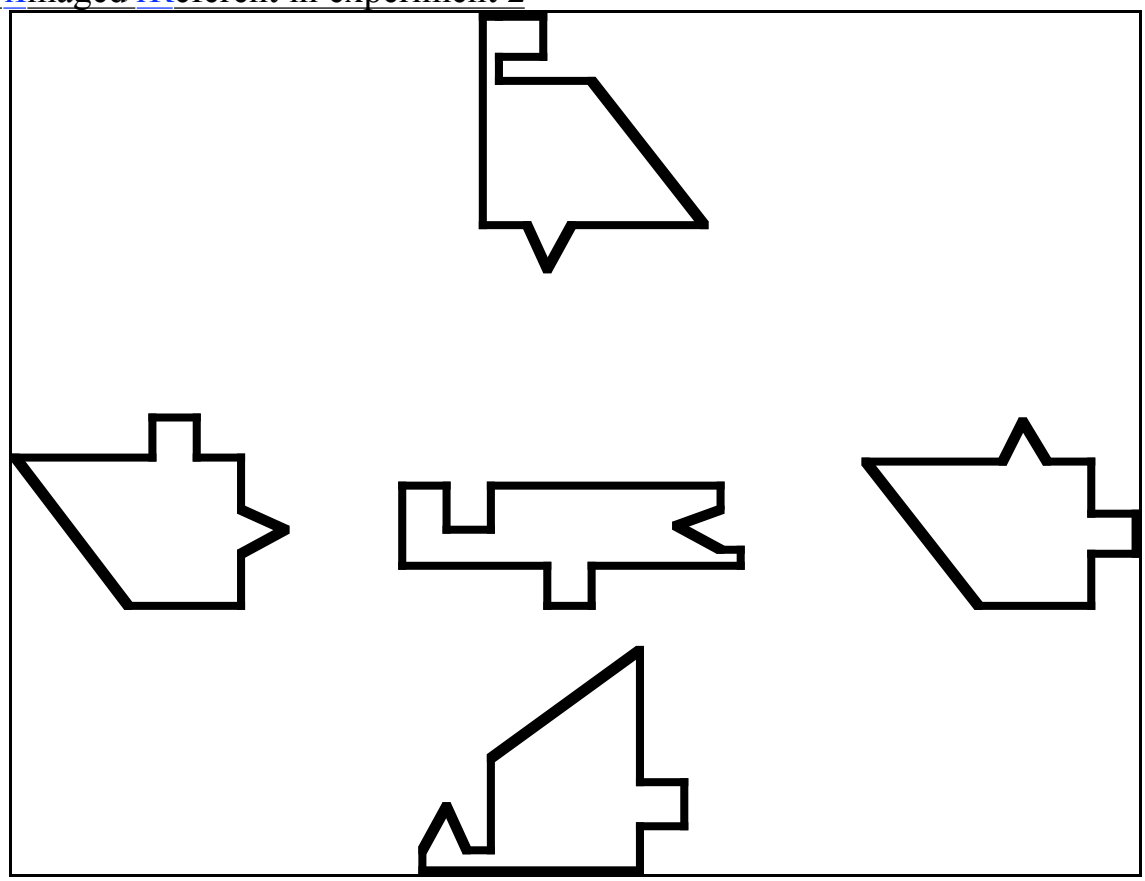

Figure 4

Duration of image consultations according to their location