Modélisation cognitive computationnelle de la recherche d'information utilisant des données oculomotrices

Francisco Lopez Orozco Lopez Orozco

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Ingénierie de la Cognition, de l'interaction, de l'Apprentissage et de la création**

Arrêté ministériel : 7 août 2006

Présentée par

Francisco LÓPEZ OROZCO

Thèse dirigée par **Anne GUÉRIN-DUGUÉ**
et codirigée par **Benoît LEMAIRE**

préparée au sein **Laboratoire de Psychologie et NeuroCognition (LPNC), CNRS UMR 5105**
et de **École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE).**

Computational Cognitive Modeling of Information Search using Eye Movement Data

Thèse soutenue publiquement le **16 juillet 2013**,
devant le jury composé de :

Monsieur, Éric Raufaste
Professeur, Université Toulouse 2-Le Mirail, Président
Monsieur, Jean-Michel Boucheix
Professuer, U. Bourgogne, Dijon, Rapporteur
Monsieur, Mathieu Roche
Maître de Conférences (HDR), U. Montpellier 2, Rapporteur
Monsieur, Éric Raufaste
Professeur, Université Toulouse 2-Le Mirail, Examinateur
Madame, Anne Guérin-Dugué
Professeur, Gipsa-lab, Grenoble, Directeur de thèse
Monsieur, Benoît Lemaire
Maître de conférences, LPNC Grenoble, Co-Directeur de thèse

Acknowledgments

Completing my PhD degree is probably the most challenging activity of my first 33 years of my life. The best and worst moments of my doctoral journey have been shared with many people. It has been a great privilege to spend my first year of my PhD in the laboratory TIMC–IMAG and the last 2.5 years in the Laboratory of Psychology and NeuroCognition (LPNC) at University Pierre Mendès France (UPMF), Grenoble, France.

I am deeply indebted to my supervisors Prof. Anne Guérin-Dugué and Dr. Benoît Lemaire. They patiently provided the vision, encouragement and advice necessary for me to proceed through the doctoral program and complete my dissertation. Both made of my stay in France a great experience. Not only their scientific criticism was important but also their everyday support. "Un grand merci" Anne et Benoît. I will have you in my mind for all the rest of my life.

Special thanks to the members of the jury: Prof. Éric Raufaste, Prof. Jean–Michel Boucheix and Dr. Mathieu Roche. I got an invaluable feedback from all of you. Your comments are helpful for the improvement of my future work.

I am also very grateful with all the members of the project ANR Gaze & EEG. With Thierry Baccino, his comments were always useful. Aurélie Campagne and Nathalie Guyader for the opportunity to run some of my experiments at the hospital under their guidance. Gelu Ionescu for the discussions about Matlab programming and data structure. Alain Chauvin, he shared with me his programs of his eyetracker experiments. These were the beginning of my eyetracking challenge. Aline Frey, all the experiments you carried out in Paris were actually part of my thesis, merci encore !!!

With regard to my family. I would like also to express all my gratitude to Sofía and Efraín who are my two best examples of hardworking people. All my love for them. Of course, to my parents who gave me the life and due their support and guidance, I am who I am. Their love provided me inspiration and they are my driving force.

Thomas and Fanny, thank you very much for your advices and help. You were there all the time ready to give me your hand. Of course, Adrián was also part of this adventure. I love all of you my dear friends and brothers. By the way, "Arriba Juchitán, señores" !!! Ociel, it is impossible to forget our bike adventures. Thanks again. Marlén and Olivier, your friendship was also a source of motivation. Merci.

Azucena, sí, usté, usté. You are also very important in my life, and especially you were a valuable source of motivation during the last 6 months of my PhD. Miss Shihuahua caray. Te quiero Azu !!!

There are many people that also played a role during this adventure but it is not easy to list all of them here.

Sincerely,

Francisco LOPEZ

Contents

Chapter 1

Introduction

1.1 Introduction

This computer science thesis is placed into the cognitive science research field and more specifically into the computational cognitive modeling or simply computational psychology.

A cognitive model is a simplified representation of cognitive and psychological processes that allows to explain the human behavior. Models are necessary because it is not possible to get a reasonable understanding of human mind only from observations of human behavior, maybe except for small and limited task domains. The problem is that the tests used to understand such mechanisms and processes could only explain superficial features of human behavior that varies along individual or groups. They are also affected by contextual factors. In simple words, it would be extremely hard to understand the human mind in this way, just like it would be extremely hard to understand a complex electronic device only by observing and testing its behavior, without any a priori ideas about the nature, the inner working, and the theoretical fundamentals of that system. Due to the complexity of the human mind and its behavioral flexibility, models are necessary to explain the details of the human mind. Without such models, experimentation may lead to the accumulation of a vast amount data without any apparent purpose.

In general, models in cognitive science may be classified into mathematical, theoretical and computational models (Sun, 2008). All of them are typically tested and validated with the cycle *Model→Hypothesis→Experiments* as is shown in Fig. 1.1. Hypothesis are formulated and they are tested and validated or not by the experiments. In the same way, from experimental observation, some working hypotheses are formulated leading to a modeling stage and ending up with new hypotheses, etc. The difference is how they are formulated. Mathematical models present relationships between variables with mathematical equations. Theoretical models describe entities, relations, and processes in rather informal natural languages. In the case of computational models the details of the involved processes are presented by using algorithmic descriptions and programs, based on computer science. Computational models provide algorithmic specificity: detailed, exactly specified, and carefully thought-out steps, arranged in precise and yet flexible sequences. Therefore, they provide both conceptual clarity and precision. The computational modeling can bring out the fine details necessary to get

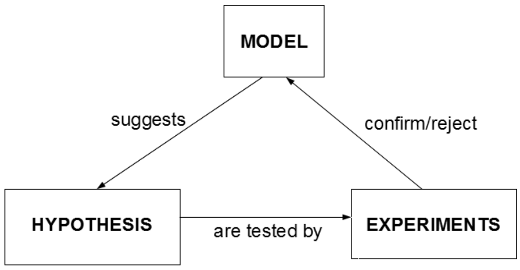


Figure 1.1: Modeling cycle for theoretical models.

an understanding of cognitive processes (Newell, 1990; Sun et al., 2005). Computational models can be run on the same stimuli people are exposed to. In other words, computational models pass the experiments.

In the current thesis, we are interested into a computational cognitive model to simulate people's behavior when they are faced to an information search task. In other words, we are interested in the category of computational models because computational modeling appears to be a powerful approach. Computational models also support practical applications of cognitive theories (Sun, 2008). With a computer simulation of cognitive processes, we would have the possibility of having a better understanding of such processes and making predictions about people's behavior faced to new situations, of information search in our case. It is important to mention that the computational models are mostly directed at answering the question of how human performance comes about, by what psychological mechanisms, processes, and knowledge structures and in what ways exactly.

We can see a model as a simulator or a generator of phenomena and data. It is a theory-building tool and not a theory itself. In particular, one may use simulations for exploring various possibilities regarding details of a cognitive process.

Very often, progresses in science are done from understanding a phenomena and making predictions of it. Then a prescription or control of such phenomena could be achieved. Computational cognitive modeling may contribute to this, because the computational model may reveal dynamic aspects of cognition giving a detailed look at their elements and their interactions during a model simulation. Such understanding may lead to uncover hypotheses or unknown aspects of cognition that may lead to predictions. Finally, prescriptions may be done if the predictions are made in reasonably accurate way.

A computational cognitive model can be integrated into a cognitive architecture. A cognitive architecture is a domain-generic computational cognitive model that capture the essential structures, mechanisms, and processes of cognition and it is used for a broad, multiple-level, multiple-domain analysis of cognition and behavior (Sun, 2008; Sun et al., 2005). They are useful and important because they provides a comprehensive initial framework for further exploration of many different domains and many different cognitive functionalities. To sum up, cognitive architectures are broad theories of cognition. In such architectures, the human cognition is represented as a structure of different components connected together. A few examples of such architectures are ACT-R (Anderson et al., 2004), EPIC (Kieras and Meyer, 1997) and SOAR (Newell, 1990). They give a baseline for studying cognitive processes for particular tasks. One of the advantage of such architectures is that all their modules have been tested and validated by different studies but typically they have a high number of parameters.

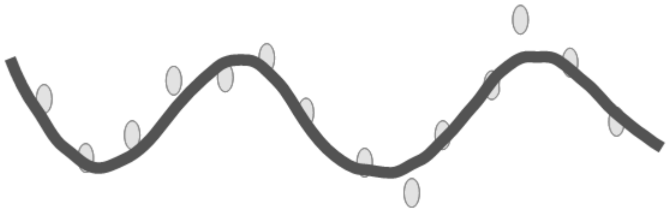
In the case of the current thesis, our models are not based on such architectures because the idea was to have a high liberty for our modeling purposes. That way, it is possible to easily separate the processes that are involved. One disadvantage of following this approach is that the models are not validated and tested by previous research works and they are more sensitive to the redundancy of parameters.

Our interest is in information search as a cognitive activity. People are used to seek information on a newspaper or a web page, composed of several blocks of text, which requires to make rapid decisions about whether to stop or not reading the current block and switch to another one. Quite often, time is a strong constraint and paragraphs are not entirely processed. People are able to judge from a couple of words whether they found a relevant text. People are concerned with this activity in everyday life and in this thesis we study such kind of information search task. More specifically, we try to simulate the eye movements during such cognitive activity at the level of fixations and saccades.[1]

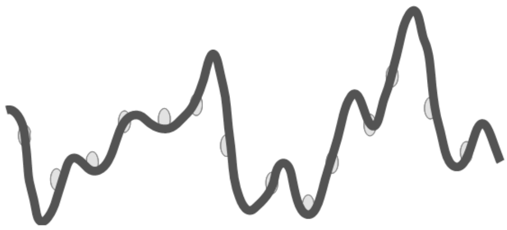
One of the objectives for the conception of our model is to build the model as simple as possible by limiting the number of parameters. The idea is to have an equilibrium between the simplicity of the model and its capacity to reproduce people's behavior. Having a model with a high number of parameters allows to reproduce the experimental data with a high accuracy. The problem here is that there is a risk of having a model *ah-doc* to this particular task and not robust enough to reproduce the data for a different task or stimuli. This model would capture noisy information in the data. In general, beyond some threshold, the closer is the model to data, the less is the prediction capacity of the model (Rissanen, 1978). For example, Fig. 1.2b shows an example of a perfect model fitting (curve) to experimental data (small spots) by using a high number of parameters. Here, it is likely that this model will not make good predictions on new data because it captures noisy information in the data. In a reverse way, a too simple model is not capable to reproduce the experimental data and it does not allow us to understand the behavior of the phenomenon studied. For this case, Fig. 1.2c shows an example of a bad model fitting the data. Then in our modeling work, we try to build a model able to reasonably replicate people's behavior with a low number of parameters

[1] A saccade is the eye movement itself which is needed to place the sight in a different place and a fixation is the period of time when the eye is practically maintained on a single location.

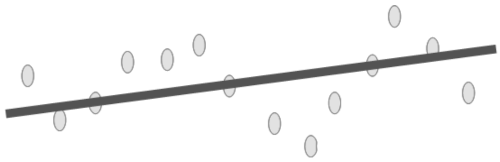
as is shown in the example of the Fig. 1.2a. We will discuss this simplicity/data fitting trade-off in the next chapters.



(a) Model fits the data well (in average) with a few parameters.



(b) Fitting with a very high number of parameters.



(c) Fitting with a very low number of parameters.

Figure 1.2: Fitting to experimental data using models with different number of parameters.

During the thesis, an experimental-based approach was followed. Psychophysical experiments with an eye-tracker[2] device were carried out in order to gather data on eye movements, create and validate the models.

1.2 Context of the thesis

This thesis is part of the research project ANR Gaze-EEG "Joint synchronous EEG signal and eye tracking processing for spatio-temporal analysis and modeling of neural activities". The aim of the project is to study the cognitive mechanisms involved during a decision-making task in an information search context. The whole project involved different research laboratories, two from Grenoble: Grenoble-Image-sPeech-Signal-Automatics Laboratory (GIPSA-Lab) & Laboratory of Psychology and NeuroCognition (LPNC) and one from Paris: Laboratory of Utilization of Technology in digital Information (LUTIN Userlab).

[2]An eye-tracker is a device that allows a real-time recording of the eye movements while people is performing a task.

1.3 Thesis organization

This thesis is organized into six chapters. Chapter 2 presents a literature review of the research done in the direction of reading and searching activity in texts. Firstly, the research works discuss both the behavioral aspects and models of reading. Secondly, research on searching on text is also presented. The chapter ends with a literature study of how decision-making processes are involved in information search tasks. Some modeling works of information search are also discussed. Chapter 3 presents the general methodology followed for the conception of our experiments, the construction, validation and improvement of the models proposed along the thesis. In Chapter 4, the first eye-tracking experiment carried out is presented. During the experiment, a very simple environment is considered when people are faced to a task of information search. The corresponding modeling work is also presented. In Chapter 5, the second eye-tracking experiment carried out is discussed. The experiment also consists of a task of information search but in a more complex environment. The corresponding modeling work is also presented. Finally, Chapter 6 synthesizes the modeling results presented along the thesis and it presents a general discussion about the limitations, drawbacks of the computational cognitive models and gives some ideas for a future work. All the textual material used during the experiments is included in the Annex 1.

Chapter 2

State of the Art

2.1 Reading as text processing

Processing a text could be done in different ways. When text processing is done as a visual comprehension of language, we can call it reading (Sereno and Rayner, 2003). This process can be also seen as a way of language acquisition, of communication and of sharing information and ideas. On last decades, many research works have been done to explain the different phenomena implied during the reading process. For example, several researchers have shown that the reading process is mainly affected by the type of task that people are performing (Carver, 1990). Among others, this effect would be present in the reading rate (number of words processed in a period of time), the sequence of words processed (if words are read serially or not), or if the words are processed more than once, etc. For example and more particularly, Carver (1990) makes a distinction between five possible kinds of processes followed by people during reading based on their reading rate. The processes are:

- *Scanning.* This process is presented when readers are looking for a particular word (600 words per minute or wpm). This is for instance what you are doing if you search for the word *phenomena* in the previous paragraph.
- *Skimming* is used when readers need to get a quick overview of the content of the text (450 wpm). This is the process you would use if you have to quickly decide whether the next paragraph is about *eye-tracking* technologies or not.
- *Rauding* is normal reading (300 wpm). This is what you do when you are reading this page.
- *Learning* is used when readers try to acquire knowledge from the text (200 wpm). This is what you would do if there were comprehension questions at the end of this chapter.
- *Memorizing* is used when readers want to memorize the text, therefore constantly verifying that information has been memorized (138 wpm).

Most of the time the *reading* process performed by people is *rauding* but is possible that they change their reading rate either up or down by moving between *scanning* (the fastest one) and *memorizing* (the slowest one). However, Carver (1990) also argued that the reading rate is not the only difference between strategies. There is also a difference between the goals, the subprocesses involved and the expected outputs according to the task that people are performing (Simola et al., 2008; Lemaire et al., 2011). We can illustrate this situation with an example. Suppose that we ask people to find the different occurrences of a particular word in a given text (e.g. find the word *market*). As we can expect, participants will more likely perform a *scanning* process. This can be seen as a *word recognition* process that require a *lexical access*. Now, in a second example, participants are asked to find two words in a text that have been transposed, such as a sentence that starts with *Market the after sweeping...*. Here, not only all the words should be lexically accessed (*word recognition*) but also the meaning of the words as they are used in the sentence should be recognised (*semantic encoding*). Then, the whole process is called *skimming* and the extra stage needed is known as *culminating* component. As expected, the resulting reading rate is slower (450 wpm) compared to the *scanning* (600 wpm). However, if people are faced with a passage where each sentence should be comprehended, things are still different because not only the same stages needed for the *skimming* (*word recognition* and *semantic encoding*) should be necessary but also each word should be integrated into the complete thought of the sentence. This stage is called *sentential integration* and the whole process is known as *rauding*. *Rauding* is normal reading. Similarly, the existence of a new component diminishes the reading rate (300 wpm) of the whole process. Now, we can imagine that people are asked to read carefully a text in order to be able to correctly answer items in a multiple-choice test about the text. Here some of the words will be read more than once with the idea to make sure that the text is being comprehended. The extra component and the whole process are called *idea remembering* and *learning* process respectively. As previously, the reading rate (200 wpm) is affected by the inclusion of the new component. Lastly, the *memorizing* process appears when people are asked to read a text with the idea to be able to recall it later. Here people will have to retell several times the sentences and this process is known as *fact rehearsal* and its latency is also observed in the reading rate (138 wpm) of the whole process.

After the explanation given so far, now it is time to recall that in the current thesis, we are interested in studying people's behavior during tasks of information search. During these kind of tasks, people are encouraged to perform a *rauding* (or *reading*) stage, in another moment they would feel more likely to follow a *skimming* strategy and seldom to perform a *scanning* step. We would argue that changing states could be observed due to different factors among others: the nature of the task and the difficulty of the text. However, before trying to get an understanding of such processes and due to the high complexity of the *rauding* process itself (also called normal reading or just reading along the current thesis), we believe in the importance of reviewing the relevant aspects of the reading activity and how they could be used as a baseline to understand more complex processes.

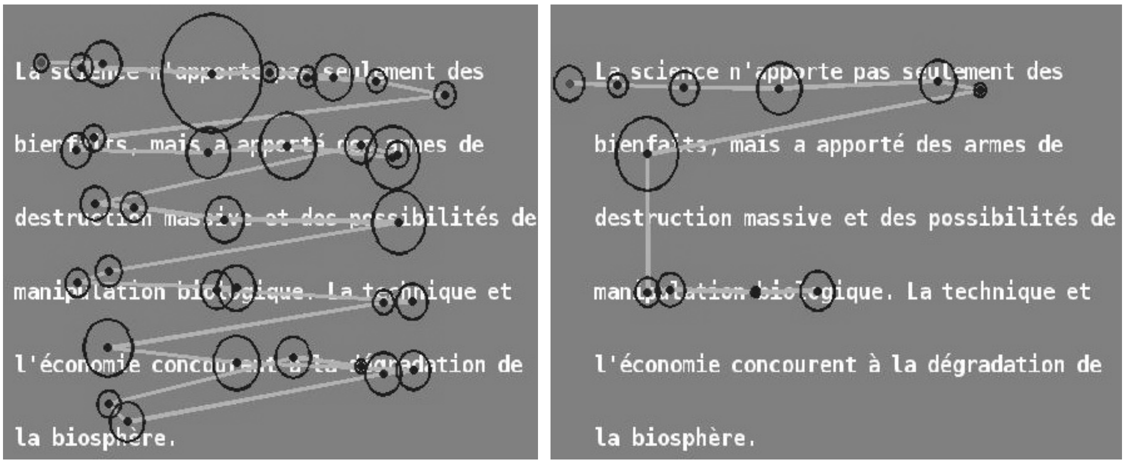
2.1.1 Reading

Reading a text is a complex task which has been widely studied in cognitive science and due to the high importance of such activity, many effort has been done to investigate the eye movement behavior during fluent reading (Rayner, 1998, 2009). Another important aspect that has favored the investigation of reading behavior is the facility of access to eyetracker. Eyetracker is a device that allows a real-time recording of the eye movements while people is performing a task. The use of the eye movement technique offers several advantages over other more traditional behavioral techniques. A central advantage is that eye movements are already a part of reading itself and people do not need to do a secondary task while reading. Most of the existing studies agree in that they have identified the oculomotor, perceptual and cognitive factors that guide the reader's decisions of *where* (space) and *when* (time) to move the eyes while processing text. They have started by making a distinction between the two basic components of eye movements: saccades and fixations.

A saccade is the eye movement itself which is needed to place the sight in a different place. It is a rapid movement during which practically no information is acquired because vision is suppressed (Matin, 1974). Saccades take time to plan (latency) and execute (duration). The codification of the target location in the visual field and initiation of the movement is made during the saccade latency and typically lasts about 175-200 ms (Rayner et al., 1983). Once a saccade is initiated, the amount of time that is takes to actually move the eyes is a function of the distance to the target. Typically, a 2-deg saccade takes about 30 ms, while a 5-deg saccade, takes around 40 to 50 ms (Rayner, 1978).

A fixation is the period of time when the eye is practically maintained on a single location (which last about 100-250 ms) and it is here that information is acquired.

Fig. 2.1 shows two experimental scanpaths (sequence of saccades and fixations of a single eye) of two different users looking at the same text. Fixations are indicated by the black spots whereas the saccades are represented by the lines connecting pairs of fixations. The scanpath pattern on the left is characterised by forward saccades moving from left to the right according to the words on each line of text, a long saccade going from the end of the current line to the first word of next line and so on until the last line of the paragraph is read. It also worth nothing that some of the words are not fixated (they are *skipped*). This does not mean that the words were not processed. In fact, as we will discuss later these words are actually processed. We can observe that the skipped words are short and not as meaningful (eg. *de*, *et*, etc.) as fixated words (eg. *science*, *technique*, etc.) which renforces the idea that short words do not need to be explicitly fixated to be processed as we will discus later. The scanpath



(a) (b)

Figure 2.1: Example of a typical scanpath a) in reading. b) in scanning.

on the left is typical of *reading* or *rauding* process. However, a different situation is observed with the scanpath to the right where there are less fixations on the text and it is clear that most of the words were not fixated. We can also notice that it seems that at the beginning (while processing the first line of the text) the participant was performing a *reading* stage (most of the words were fixated serially), but he switched to a different strategy by fixating only one word on the second line, by skipping the third line and by doing some fixations on the line number 4. The scanpath is typical of a *scanning* process. From this example, we can notice that participants can follow different strategies for reading a paragraph and that a same participant can switch to a different way of processing the paragraph that he is looking at.

According to Sereno & Rayner (2003) the reading process is more than only identify individual words of a text. Several cognitive processes are involved during the course of reading. Figure 2.2 illustrates (see Sereno & Rayner (2003) for further details) the time course of events occurring during a single fixation in reading. Once a word is fixated, the information acquired from such a fixation spent about 60 ms to be available in the visual cortex where the lexical processing begins. Then the fixation duration varies according to the lexical difficulty of the word (Rayner and Pollatsek, 1989; Rayner, 1998). For example, frequent and predictable words from context are fixated for less time than those that are not, but in general the word is processed within the first 100-200 ms of the fixation because the next eye movement should be also programmed. Then a signal is sent to eye muscles to start the saccade. The time when the word recognition is done does not correspond exactly with the fixation time on this particular word. Several combined approaches have tried to explain the different stages of processing

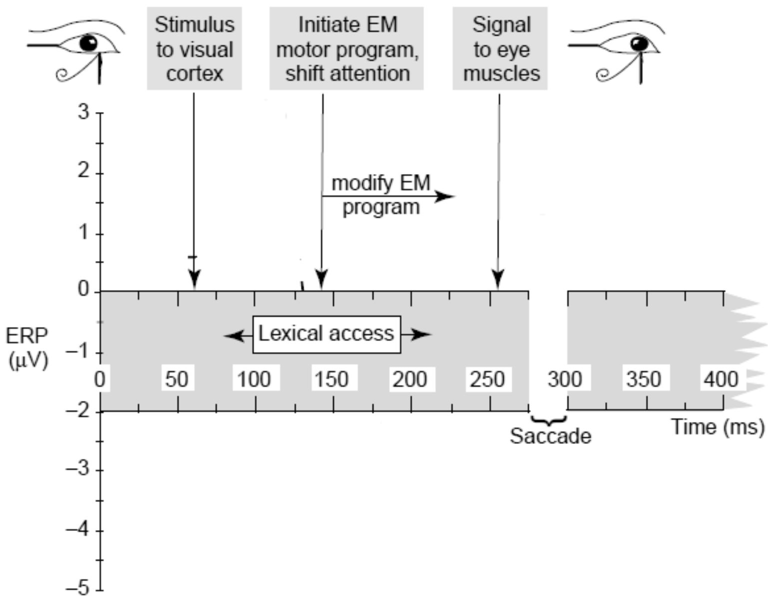


Figure 2.2: Time line of processing a word in reading. After the saccade from the word, a new fixation begins. Adapted scheme from Sereno & Rayner (2003).

within a fixation. For example, eye movements recording technique is combined with the registration of the brain activity during cognitive tasks (eg. reading, searching, etc.). However, despite the vast effort done in such direction, what drives the eyes forward in text has not been fully resolved.

In a reading process, the usual distribution of fixation duration follows approximatively a Gaussian distribution with a mean around 150-200 ms. For example, Fig. 2.3 shows a typical distribution fixation duration (in ms) and saccade length (in letters) for two particular experiments: reading a normal word text vs reading the corresponding z-string text. In a z-string text, all the letters are replaced with the letter z, preserving inter-word spaces, punctuation, and letter cases (Nuthmann et al., 2007). People use z-strings to remove the semantic part of reading. Fixation durations (left plot) are significantly longer in z-string than in normal reading and it is observed by the right shift of the fixation duration distribution for z-strings. This supports the finding that participants produced longer fixation durations when engaged in mindless reading (Rayner and Fischer, 1996). However, the mean length of forward saccades did not differ between the two conditions (right plot), but a significant difference is observed between the mean length of regressive saccades. Given a fixation point, a forward saccade is done when the second fixation causes an eye movement to the right of the first fixation (positive) and a backward saccade is done when the second fixation causes an eye movement to the left of the first fixation (negative). People perform regressions

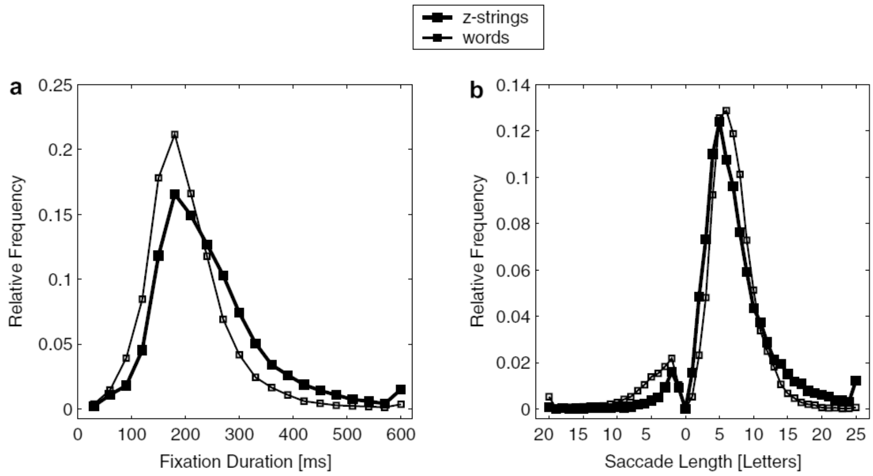


Figure 2.3: (a) Distribution of all observed fixation durations during z-string reading (full squares) vs. normal reading (open squares). (b) Distribution of all observed saccade lengths. Distributions taken from experiments described in Nuthmann et al. (2007).

(negative saccade lengths) less frequently during reading of z-strings than during normal reading. Regressive saccades are much less useful (all the letters are the same) during the z-string reading than in normal reading. The idea behind this example was to claim again the fact that there is a task effect over the way people process a text. For example, if people were also asked to read the normal text aloud, this would result in a fixation duration still longer because in that case participant will be performing two different tasks at the same time. However, with respect to the length of saccade, a significant difference would not be expected.

We have roughly explained how and when the two basic phase of eye movements (fixations and saccades) are generated and also how they are modified by factors like the task type that people are performing. Now, we will discuss about one of the most important physical characteristic of our vision system to understand its effect in the way people process a text during reading.

The retina is the place where the visual stimuli that come from the real world are formed. Then, they are transmitted to the visual cortex. The anatomy of the retina is characterized by a layered structure where the cell density of each layer decreases when the eccentricity increases. The center of the retina aligned with the optical axis is called fovea. Here the density of the photoreceptors is maximum, and thus also the visual acuity. Farther from the optical axis is located the parafoveal area and thereafter the peripheral area, with an increasing of the eccentricity and the distance from the foveal area. There the density of the photoreceptors decreases and also the visual acuity.

To sum up, acuity is very good in the fovea, it is not so good in the parafovea, and it is even poorer in the periphery. The distinction about the three areas is observed during the reading process.

Figure 2.4 shows a simulation [1] about how the visual acuity is present during reading a text. Only in the fixation point (fixation on the word *four*) the visual acuity is of 100% (at the fovea). The degradation of vision (at the parafovea and periphery) is observed both to the right and to the left of the fixation point as illustrated in the Figure 2.4. Because of the anatomy of the retina and some limitations due to the fact that only a part of the retina has 100% of visual acuity, eye movements are needed. Viewers move

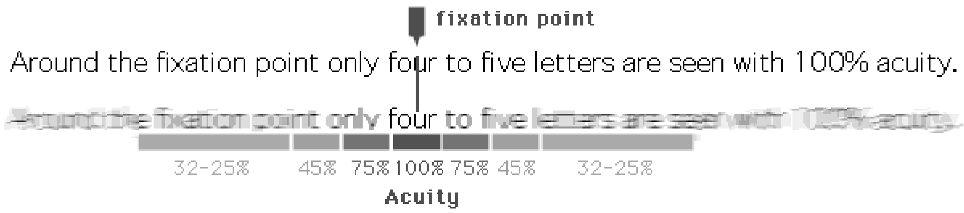


Figure 2.4: Simulation of the visual acuity during a reading process (Hunziker, 2006). Fixation is on the word *four*. The degradation of vision is observed around this point (both to the left and to the right of the fixation point).

their eyes so as to place the fovea on that part of the stimulus they want to see clearly. Then, as we have also seen that the human eye has sharp enough vision to read text in a well limited area, it is important to know some details about the factors that could influence the size-geometry of such area. This area is known as the visual or *perceptual span* of effective vision and a short description is given below.

2.1.1.1 Perceptual span in reading

The amount of information available to process and to use during each fixation is limited by an asymmetric window extending roughly 3-4 character spaces to the left of fixation to about 14-15 character spaces to the right of fixation (Rayner et al., 1980, 1982; Underwood and McConkie, 1983). However, Rayner (1998) also showed that the area from which a word can be identified extends to no more than 4 characters to the left and no more than 7-8 characters to the right of fixation and corresponds to the word identification span which is still more limited than the general perceptual span. Moreover, Pollatsek et al. (1993) showed that even if information of the next line is processed during a reading task, participants are not capable of getting some semantic information. With these results, we can imagine that there is not a one-to-one mapping between each fixation and the processed words. In fact, during each fixation, not only

[1] http://www.learningsystems.ch/multimedia/vis_e02.htm

the fixated word is processed but it also possible to get a preview of the word located to the right of the current fixation. When another word or a nonword initially occupies the target location, this preview is not valid. Some research works have revealed that when readers have a valid preview of the word to the right of fixation, they spend less time fixating that word than when they do not have a valid preview. The effect is called preview benefit in reading and its description is given now.

2.1.1.2 Preview benefit in reading

The preview benefit effect in reading was found by using a boundary paradigm (Rayner, 1975). In the boundary paradigm, an invisible boundary is defined just to the left of a target word. Before a reader crosses the boundary, there is typically a preview different from the target word but when the eyes cross the boundary, the information of the target word is in the preview. The size of this preview benefit is typically of the order of 30-50 ms. The exact amount of preview benefit is modulated by the difficulty of the current fixated word: the more difficulty to process the fixated word, the less is the preview benefit obtained and viceversa. Another interesting result is that the preview benefit is larger within words than across words.

An investigation about the characteristics of the word to the right of fixation than can influence the duration of the current fixation has been motivated by the preview benefit effect. This is largely known in the literature as the parafoveal-on-foveal effect. It is still in a controversial discussion and given the possibility of mislocalized fixations and the fact that most of the positive evidence for these effects is based on corpus-based analyses, it seems reasonable to view such effects with caution according to Rayner et al. (2006).

After having presented some of the limitations but also positive aspects of the vision system that have a direct influence over the reading process, it is time to look at the mechanisms that govern eye movements.

2.1.1.3 Control of eye movements in reading

The two interesting questions concerning eye movements are: what determines *when* to move the eyes, and what determines *where* to move them? It has generally been considered that these two decisions are made somewhat independently. Rayner & Pollatsek (1981) showed in a experiment with some manipulations that saccade length and fixation duration are affected independently, reinforcing the view that the decisions about *where* and *when* to move are made somewhat independently.

At least in languages like English and other alphabetic languages, the decision of *where* to move the eyes next is influenced by two main factors: length of words and space information. The length of the fixated word and the corresponding length of the word to the right determines the length of the next saccade (Rayner, 1979), because if the word to the right of the current fixation is either very long or very short, the saccade will tend to be longer than when a medium-size word is to the right of fixation. A second factor that have some influence in *where* to land the next saccade is the space between words because some experiments have shown that when spaces are removed, reading slows down by as much as 30-50%. Contrarely, other experiments have demonstrated that when space information was provided for readers of Thai (who are not used to reading with spaces between words), they read more effectively than normal (Kohsom and Gobet, 1997). Surprisingly or not, semantic preprocessing of words does not influence *where* readers move their eyes (Rayner, 2009).

Another factor that has some influence in *where* to move the eyes is the landing position effect. Rayner (1979) demonstrated that readers' eyes tend to land between the beginning and the middle of the word. This position is known as the preferred viewing location (PVL). When readers' eyes land at a nonoptimal position in a word, they are more likely to refixate that word with a very short saccade (Rayner et al., 1996). At the end, the final landing position also depends on the prior launch site. There is other factor which influences the landing position and it is related to the fixation location on a word. Recognition time is minimized when the fixation is at the centre of the word. This is called the optimal viewing position (OVP). Moreover, when readers make only one fixation on a word, if that fixation is at the centre of the word, then the fixation is longer than when a single fixation is at the end of the word. This is called the inverted optimal viewing position (IOVP) effect. Naturally, during the reading process some words are not fixated. This is called the skipping effect. A word skipped does not mean that the word was not processed. In fact, the word is processed in parafoveal vision before or after of skipping the word because the context helps to identify that word. The most important factor that influences the skipping effect is the word length (Rayner, 1998) because in general short words are much more likely to be skipped than long words. Skipping words are also regulated by the contextual constraints. Words that are highly predictable by the context are much more likely to be skipped than those that are not predictable.

Concerning the second question about *when* to move the eyes it has been clear that ease or difficulty of the word processed has a strong influence. Basically, all the variables that influence *when* to move the eyes can be grouped in lexical and linguistic variables. Some of them are word frequency, word predictability, number of meanings, phonological properties of words, semantic relations between fixated words and prior words, word familiarity among others (Rayner, 2009). It has also been demonstrated that morphological properties of the words present some influence over the fixation times on target words. For example, in longer compound words, a high frequency beginning or ending lexeme is fixated shorter than low-frequency lexemes. To sum up, the word frequency, age of acquisition and the word predictability are the variables with a strong and immediate effect on how long readers look at word and the other variables has an influence on how soon readers move on in the text.

In any case, there is no doubt that cognitive processing activities have a strong influence on *when* the eyes move.

Finally and as it was already mentioned before, the main objective of the current thesis is to develop a computational cognitive model of people looking for information. We have started by reviewing some aspects of the behavior during the reading process, we should now take a look into models of reading before to present the cognitive task in which we are interested the most: searching for information.

2.1.2 Models of Reading

Two of the most representative computational models for reading will be presented in this section: E-Z reader and SWIFT. E-Z reader corresponds to the category of models where words are serially treated: only one word is processed at a time, because the next word could be processed even if the eye is still in the current word. This is because attention and eye movement do not progress together in E-Z reader. In the case of SWIFT model, the words are treated in parallel: more than one word could be processed at a time.

2.1.2.1 E-Z Reader

The E-Z reader model (Reichle et al., 1998, 2003, 2006) describes a theoretical framework to understand how word identification, visual processing, attention, and oculomotor control determine *when* and *where* the eyes move during reading. The model operation is based on two important assumptions: the attention that is necessary to process and identify printed words during reading is done serially (one word at a time) even if this is opposed to what is assumed in most alternative models of eye-movement control in reading, which assume that attention is done in parallel to support the simultaneous processing of multiple words (Engbert et al., 2005; Reilly and Radach, 2006) or that attention has little or nothing to do with eye-movement control. A second important assumption of the model refers to a decoupling of two signals: the signal to shift covert attention to the next word and the signal to make an eye movement to the next word. The recognition phase of a $word_n$ is composed of three different stages:

- V: A visual processing in a preattentive manner. Low-spatial frequency information about word boundaries is used to select the upcoming saccade target and high-spatial frequency information about letter and/or word features is used to identify the individual letters and words.

- L1: Familiarity check (in an attentive manner) causes the eyes to prepare to move forward to the word during reading. This is attenuated by visual acuity, as

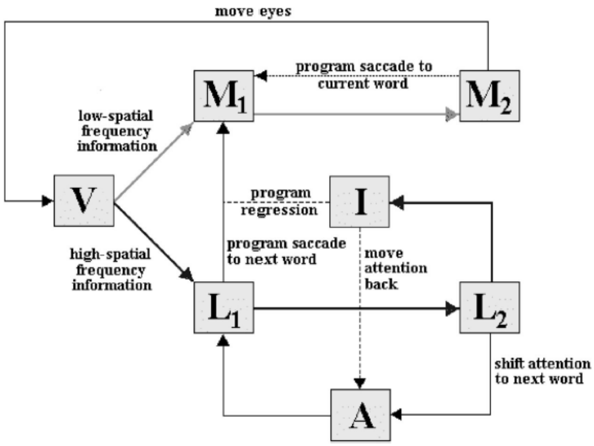


Figure 2.5: Schematic diagram of the E-Z reader model of eye-movement control in reading (Reichle et al., 1998, 2003, 2006). The block components are: preattentive visual processing (V), familiarity check (L1), lexical access (L2), attention shift (A), postlexical integration (I), labile saccadic programming (M1) and nonlabile saccadic programming (M2).

a function of foveal eccentricity, or the mean absolute distance in character spaces between each of the letters of $word_n$ and the current fixation location.

- L2: Lexical access (in an attentive manner) is the trigger to shift attention to the next word. This reflects whatever minimal amount of time is necessary to activate the word's meaning.

Then, two processes are simultaneously executed. First, attention shifts (A) from $word_n$ to $word_{n+1}$. Second, the meaning of the word is subjected to some minimal amount of higher level language processing called postlexical language processing *integration* (I). The saccadic programming in E-Z reader is completed in two stages: a preliminary labile stage (M1) that can be canceled by the initiation of subsequent saccadic programs, followed by a nonlabile stage (M2) that is not subject to cancellation. Fig. 2.5 shows a schematic diagram of the E-Z reader model with its block components. In E-Z reader, a saccade is posited to be directed toward the optimal viewing position (the center of the word), but saccades usually land at a location other than the intended target because of both systematic and random error.

The lengths of the executed saccades (in character spaces) are therefore the sum of three components: the intended saccade length, systematic error, and random error. The systematic error (on average) causes short saccades to overshoot their intended targets and long saccades to undershoot their targets. This error is modulated by the starting fixation duration: the longer the fixation duration, the higher the accuracy of

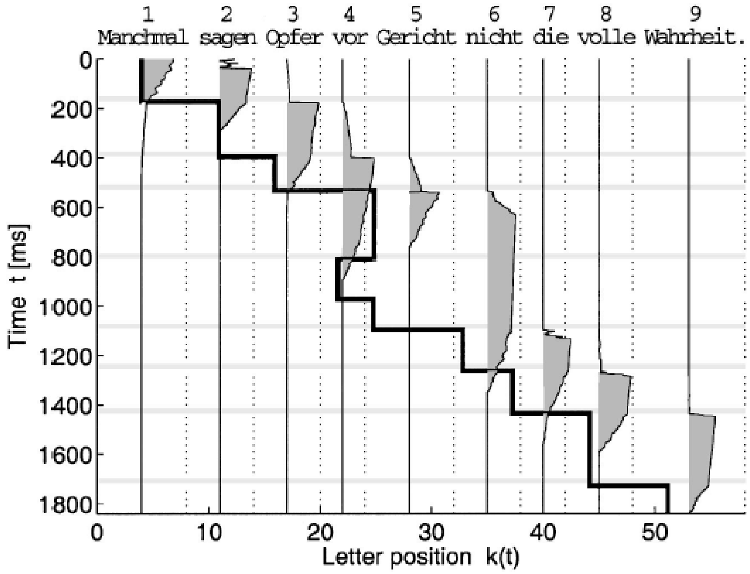


Figure 2.6: Example of the numerical simulation of the SWIFT model (Engbert et al., 2005). Fixation position $k(t)$ is indicated by the dark black line (in units of letters), activations $a_n(t)$ by the gray areas and saccades by the lighter horizontal lines.

the saccade. The random error increases with the amplitude of the saccade.

The new location of the eye is propagated from the retina to the brain so that it can be used to the lexical processing, which continues using information from the previous viewing location during the saccade and for the duration of the eye-mind lag. This stage is assumed to be preattentive, with low-spatial frequency information about word boundaries being used by the oculomotor system to select the upcoming saccade target and high-spatial frequency information about letter and/or word features being used to identify the individual letters and words.

So far, a quick review about the E-Z reader model was given. The model is still valid and on last years has been used to explain also many phenomena related to non-reading tasks (Reichle et al., 2012).

2.1.2.2 SWIFT

Engbert et al. (2005) presents SWIFT-II (an advanced version of SWIFT-I, (Engbert et al., 2002)) that is a model with a strong mathematical background which contributes to the understanding of the control of eye movements during reading. The main difference between SWIFT and the EZ-reader model is that in the first one more than one word can be seen/processed in parallel way. SWIFT is based on the following principles:

- PRINCIPLE 1: The processing of an activation field is spatially distributed: the target word to be fixated is selected from a competition between words according to their activation. Activation can be seen as a saliency map of words as is indicated (in gray) in Fig. 2.6. Due to the fact that activations are done in a parallel way over several words, the processing is also distributed among several words at a time.

- PRINCIPLE 2: Pathways for saccade timing and saccade target selection are treated separately: problems of *when* to start the next saccade program and *where* to go next, are not coupled.

- PRINCIPLE 3: Autonomous generation of the saccade and foveal inhibition: fixation duration depends on the programming time for saccade and it is modulated by a foveal inhibition process that increases with difficult words.

- PRINCIPLE 4: A saccade programming is done in two stages: a saccade is programmed during the labile stage, but a new saccade could be programmed with the cancellation of the first saccade. In the nonlabile stage, the target is selected and the saccade cannot be cancelled.

- PRINCIPLE 5: Inherently, there are systematic and random errors in the length of saccades.

- PRINCIPLE 6: Mislocated fixations are corrected: when a saccade land on unintended words, a new saccade program starts immediately.

- PRINCIPLE 7: Saccade latency is modulated by the amplitude of the intended saccade: the time needed to start the saccade depends on the amplitude of the saccade to execute.

Fig. 2.7 illustrates a general diagram of the SWIFT model. By using PRINCIPLE 1, word recognition is implemented as the spatially distributed process mentioned previously. According to PRINCIPLE 2, the two separate pathways used by the foveal inhibition determine the saccade target selection (*where*) and the saccade timing (*when*). The lexical decision circuit can influence saccade timing by foveal inhibition only with a time delay τ (PRINCIPLE 3). As it was also explained before, saccade programming is splitted in two different states (PRINCIPLE 4). The nonlabile stage (a point of no return) is preceded by a labile stage (a point of return) and target selection occurs at the transition from labile to nonlabile stage.

The estimation of the model parameters is based on word measures. The measures were:

- Fixation durations: first fixation duration, second fixation duration, single fixation duration, and total reading time and their corresponding distributions.

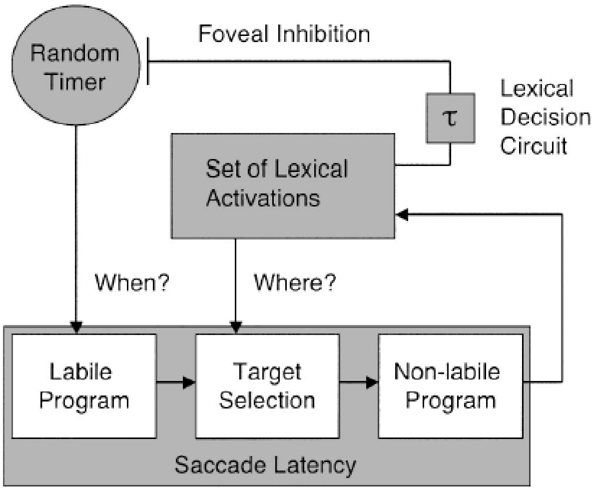


Figure 2.7: Diagram of the SWIFT model (Engbert et al., 2005).

- Fixation probabilities: skipping probability, probability for two fixations, probability for three or more fixations, and regression probability.
- Relations between fixation duration and within-word fixation duration for first and single fixation durations.

All the detailed fitting procedure is explained in Appendix A of Engbert et al. (2005). The sum of mean squared normalized errors of fixation durations and fixation probabilities per word are the measures of the model performance.

Let us summarize how the model works: given a new phrase, the model simulates reading by indicating *when* and *where* to fixate. A typical example of this procedure is illustrated in Figure 2.6 by plotting the time evolution of the set of activations $a_n(t)$ and the fixation position $k(t)$ along the vertical axis. The sequence of words fixated in this example is $\{1, 2, 3, 5, 4, 5, 6, 6, 8, 9\}$. From this numerical example, we can see some of the typical phenomena of reading as *word skipping* (eg. words 4 and 7 are skipped in first-pass reading) and this mechanism is observed by the parallel activation of several words in the example. Some words are *refixated* (eg. word 6). Also regressive saccades are present (word 4 was skipped in the first pass and later fixated with a regression). The other examples shown in the figure can be analyzed in a similar way.

Without debt, the SWIFT model is a plausible computational model that represents a framework that encompasses almost all types of eye movements observed during the reading processes: forward saccades, refixations, word skippings, and regressions. For that reason, with the SWIFT model it is possible to reproduce a number of well-known measures of eye-movement control during reading like average fixation durations and fixation probabilities, distributions of within-word landing positions, and interword regressions.

Now that we have reviewed in a very simplified way the reading process and two of the most representative models of reading, we are going to look at the type of cognitive task we are interested in the current thesis: information search on textual material.

2.2 Searching on texts

In this section, we would like to make the distinction between reading and non-reading tasks on textual material. More particularly, we will focus on one of the well-known non-reading tasks: searching on text.

2.2.1 Searching

A searching task on texts could be as simple as looking-up for a particular word (as it is done in a dictionary), looking for the most relevant article related to a given subject, looking for specific information to answer a particular data question, etc. All these activities are different from pure reading, because all of them imply to manage at the same time the text processed so far but also the search goal in mind. Very often, information search activity implies making decisions because information search is a process in which a person seeks data or knowledge about a problem, situation, or artifact. Thus, people need to make *semantic judgments* in such tasks.

Since that our ultimate goal in the current research is to build a model to make predictions about how people look for information on textual material, some of the most representative works already done in information search area are now presented.

2.2.2 Models for searching

In this section, we will concentrate on the description of some representative modeling works done for non-reading tasks and more specifically for searching tasks. E-Z reader model for non-reading tasks is presented, and a hidden Markov model learned from eye movements for information search tasks is also described. Next, a very simple model based on different visual strategies of information search is presented. At the end, a computational cognitive model of information search is also detailed.

On last decades there has been considerable progress in understanding such tasks but the progress has been slower at least because the demands of these tasks are much less constrained than those of reading (Reichle et al., 2012). For example, in a scene viewing task, eye movements reflect both the variability of the physical layout of scenes and the goals of the viewer. There are a very limited number of computational models of eye-movement control in non-reading tasks due to its complexity among other reasons. Moreover, most of them account for only limited aspects of the tasks being simulated. For example, they simulate *when* or *where* the eyes move during scene viewing but not both. They have also been developed and evaluated only within each of their respective task, with any consideration of whether the theoretical assumptions of the models generalize to other tasks. A well-known exception for this is the modified version of the E-Z reader model (Reichle et al., 2012) since as it was presented earlier, the original conception of such model was for reading purposes but a modified version of the model for non-reading tasks is presented now.

2.2.2.1 E-Z Reader for non-reading tasks

Reichle et al. (2012) present the E-Z reader model to simulate eye movements in non-reading tasks as a unified framework for understanding the eye-mind link, as opposed to most of the research done using that model.

They use variants of the E-Z reader model of eye-movement control in reading to simulate eye-movement behavior in several of non-reading tasks. Their strategy in the simulations was to first examine whether the basic framework of the E-Z reader model could be used to simulate eye movements in three non-reading tasks: (a) visual search for a specific target word, (b) z-string "reading" and (c) visual search for an *O* in clusters of Landolt *Cs* (rings with a small segment missing, making them look like *Cs*).

With the simulations, they demonstrated which of the assumptions are sufficient to account for the similarities/differences in eye movement patterns in both, reading and non-reading tasks. In the target-word search, subjects are instructed to scan sentences for a particular target word (e.g. looking-up the word *book*) and to indicate all occurrences of the target word with button presses. In the z-string "reading", subjects are instructed to "pretend" that they are reading sentences in which all of the letters in each of the sentences have been replaced by the letter *z*, but with capitalization, spaces, and punctuation preserved (e.g., the sentence "During the year, many questions were made" would thus be converted to "Zzzzzz zzz zzzz, zzzz zzzzzzzzz zzzz zzzz"). In the Landolt-*C* task, subjects are instructed to search through an array of Landolt *C*s and to indicate whether there was also a letter *O* in the display with a button press. The "gap" size (measured in terms of number of pixels) in the Landolt *Cs* was varied, making the task of discriminating them from the target *O*s more or less difficult. All these non-reading tasks are similar because they logically require little or no language processing (e.g., practically everyone without any particular level of knowledge can scan through sentences and decide whether word *engine* is present or not, by performing a pattern matching process).

They assumed that attention is also allocated serially (as in E-Z reader model for reading task) and simplified the modeling work as a basis for generating interesting hypothesis. Due to the fact that the non-reading tasks involved very little of high level language cognitive processing, the E-Z reader model's assumptions about how higher level processing influences progressive eye movements was disabled to analyze how object processing and identification affect the patterns of the eye movements for the first-pass when there is no any assumption about higher level processing.

Those assumptions related to the programming saccadic and its execution can be generalized to a wide variety of visual-cognitive tasks, because in fact they were derived from reading and non-reading tasks for the E-Z reader model. However, for example it is expected to observe a variability for the precise metrics of the saccades because the *optimal* saccade length depends on the nature of the task. Lastly and concerning the assumptions referring to the visual processing, they also have considered that the degree to which limitations in visual acuity influence the identification of different visual stimuli may vary in function of the nature of the task.

An interesting result is that, although most of the basic principles of the E-Z reader model are applicable to other visual cognitive tasks, the assumption that the familiarity check (L1) (See Fig. 2.5) to initiate a saccadic programming is a preliminary stage of word identification is may be limited to reading or reading-like tasks. A second one refers to the fact that the overall time that is required to trigger the saccade can vary and can be modulated by cognitive variables (eg. word frequency). However, the authors acknowledged that these results should be seen as working hypotheses more than true conjectures because they do not provide definitive proof that they are true.

All the simulations also demonstrated that a single computational framework is sufficient to simulate eye movements in both reading and non-reading tasks but also suggest that there are task-specific differences in both saccadic targeting (*where* to move the eyes?) and the coupling between saccadic programming and the movement of attention (*when* to move the eyes?). The architecture of the E-Z reader model that guides eye movements in tasks such as reading is flexible enough to accommodate tasks other than reading. These results suggest that some aspects of the eye-mind link are flexible and can be configured in a manner that supports efficient task performance.

Now a hidden Markov model learned from eye movements for information search tasks is presented.

2.2.2.2 Markov model for processing cognitive states from eye movements in information search

Simola, et al. (2008) present a modeling work using a discriminative hidden Markov model (dHMM) whose parameters were learned from experimental eye movement data. With the proposed model, they are able to make predictions about how processing states alternate (they assumed variations of the eye movements pattern within a task) when people is performing an information search task. For example, at the beginning of a information search task people could perform a *scanning* strategy characterised by rather longer saccades without any preference on direction and fewer backward saccades. Eventually they could switch to a *reading* strategy with frequent short forward saccades with an average fixation duration about 200ms, etc. This is the same problem that we have discussed around Fig. 2.1a and 2.1b (p. 10), where we showed how participants can follow different strategies for reading an item of information and how they can switch to a different way of processing the item that they are looking at.

Simola, et al. (2008) followed a reverse inference approach by trying to infer the hidden cognitive states from an observable eye movement behavior. The information search tasks were: a) looking-up a word (W), b) answering a question (A) and c) finding subjectively the most interesting topic from among various ones (I). The association between eye movements and language processing is made by modeling the sequences of fixations and saccades and by assuming the existence of states as indicators of the cognitive system that alternates between different states of processing. The sequence of fixations and sacaddes were treated as time series and modeled by the dHMM. To learn the model, four features were extracted from each fixation-saccade data:

1. Logarithm of fixation duration (in ms).
2. Logarithm of outgoing saccade length (in pixels).
3. Outgoing saccade direction (four different directions) + a fifth state to indicate the trial end.
4. A flag variable to indicate if the currently fixated word was previously fixated.

Three states were uncovered and designated by comparing the parameter values to literature on cognitive tasks (eg. on reading) to describe the processing behavior reflected in the eye movement data. This was possible due to the consistency observed into the parameter values of the dHMM in the three states for the three different tasks (see Simola, et al. (2008) for details). For example, at the beginning of the task and with a probability of 67% (32+20+15) participants performed a behavior termed *scanning*. Fixation duration were short (135 ms) which agrees with the fact that shorter fixations are associated with easier task (Rayner, 1998). With a probability of 43% (16+20+7), the second state observed is labeled as *reading*. Here the percentage of regressive saccades correspond to the previous findings suggesting that in normal reading about 10-15% of saccades are regressions (Rayner, 1998). Also, the average saccade length was 10.3-10.7 letters, which corresponds to the average length of a word (9.9 characters), plus a space between words. Finally, participants turned into a stated termed as *decision* because the features indicated a lot of rereading of the previously seen lines. This states is characterised by the high frequency of forward and backward (20-30%) saccades. Saccade lengths (10.7 letters) also correspond to the length of a word, and occurred within the same line (with 75% probability). With a high probability of 78-86%, new fixations were in a previously fixated words and were shorter (175 ms) than in the reading state.

A schematic diagram of the dHMM that best fitted the data is shown in Fig. 2.8. The diagram indicates the probability transitions between the states for each of the

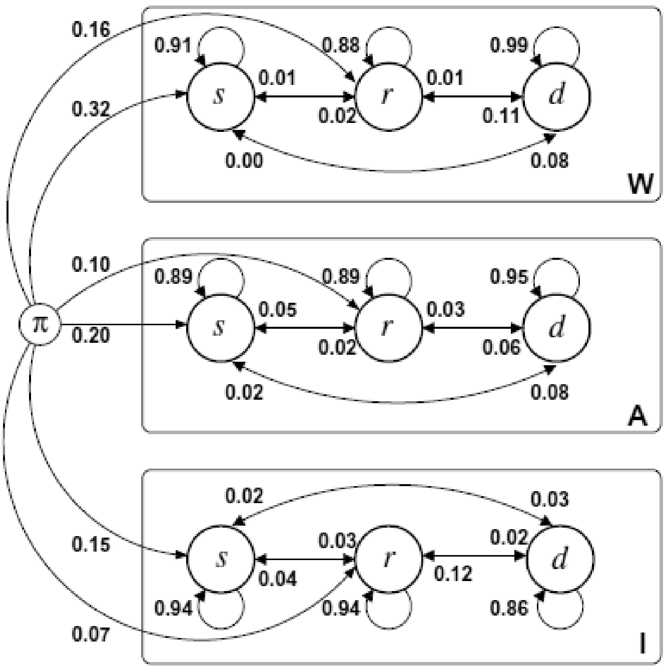


Figure 2.8: Transition probabilities and topology of the dHMM (Simola et al., 2008). Symbols used are circles (hidden states), arrows (transitions) and their probability values. Tasks: W (looking-up a word), A (answering a question), I (finding the most interesting topic), s(scanning), r(reading) and d(decision).

experimental condition (W,A,I). The beginning of a sequence is indicated by π and the arrows are the state transitions. For example, we can notice that in the W and A conditions, participants switched more often from scanning to decision (with 80% probability) than to reading (20% probability). This can be seen from Fig. 2.8 by comparing the associated transition probabilities (8% vs. 2%). This is because both task are relatively easier than in the I condition where the transition probabilities to switch from scanning to decision/reading were the same (3% probability). We can also observe that in W condition, transitions from *decision* state occurs in rare cases (1% probability) whereas in the A condition, these transitions occur more often (5% probability) and in the I condition they are still more frequent (14% probability). As previously, we can explain this by the different grade of difficulty among the tasks. A possible explanation is that participants could have adjusted their processing states according to the current task demands as is proposed by Carver (1990).

We can imagine several implications from the proposed model. Interactive information search applications could benefit from it, because they could learn and adapt themselves to the goals and intentions of the users. On the other hand new experiments might provide valuable information about the link between the processing states observed and the eye movement patterns by combining eyetracking and other tecnologies used in cognitive science research (eg. electroencephalography, fMRI, etc.). Lastly and as the authors also mentioned, it will be also interesting to study how the processing states would be generalizable to other cognitive tasks (those different from search tasks) and if people differ in the way they switch between processing states.

2.2.2.3 An information search model based on visual strategies

Lemaire et al. (2011) investigate visual strategies of eye movements in processing textual material and proposed a 5-parameter cognitive computational model which simulates the people's oculomotor behavior considering the differences in such visual strategies. As in Simola et al. (2008) and Cole et al. (2011), they also showed that there is task effect on the processing patterns (visual strategies) during information search. Then, their idea was to build a model that would account for these various visual strategies. The different processing patterns they followed are described by Carver (1990). The processing patterns differ in reading rates, the length of saccades, fixation duration and number of regressions.

The aim of the modeling work was to build a general model by adjusting the parameters according to a one of three possible tasks: a) searching a target word, b) searching the most relevant paragraph according to a goal and c) memorizing a paragraph. Basically, the model selects which word to fixate next among all words in a paragraph, by computing it with the 5-parameter mathematical expression of the model. The overall process is an iteration of two steps: weighting all words and selecting the best weighted

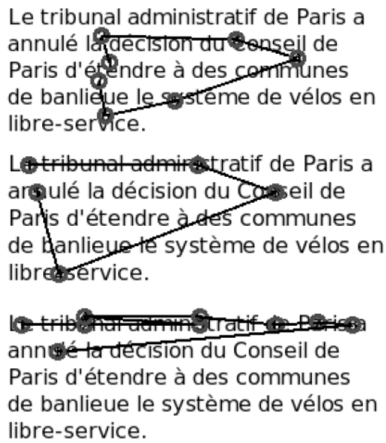


Figure 2.9: A participant scanpath (top). Two artificial scanpaths (middle and bottom), Lemaire, et al. (2011).

one. The five variables of the model are: 1) word length, 2) distance to the current fixation, 3) saccade horizontality, 4) shape similarity with the target and 5) newness of the word. Then, given a task (eg. looking up a target word, memorizing a paragraph, etc.) the model assigns predefined weights to variables. Then each word is given a weight by simply combining the values given by all variables. The variables were weighed differently according to the task involved (they play a different level of role among the tasks). These weight were learned from the data. For example, the length of words plays an important role in the task b), a reduced role in the task a) and no role at all in the task c). However minimizing the distance to the current fixation is crucial in b), not so important in scanning a) and slightly necessary in c). Also the horizontality of saccades is very important in c) and b) but not much in a). Finally visual similarity of word shape is only necessary in a). The simulation results of the model are close to those observed in the participant's data and they are achieved with a simple linear combination of the five parameters of the model. For example, Fig. 2.9 shows an experimental participant's scanpath (top) and artificial scanpaths generated by the model with different parameter values (at middle and bottom). The middle scanpath is close to the participant's scanpath and it is more likely that both are following a scanning strategy instead of a reading as is observed in the artificial scanpath at the bottom. The idea of the example is to show how it is possible to generate different scanpaths by using the same model but setting its parameter values differently.

The fastest participant performance (less fixations) was observed in task a) because it is likely that users tend to prefer words which match with the target. Task b) is particularly of our interest for the purposes of current thesis. However, we thought that the inclusion of a semantic component may be particularly interesting because by this

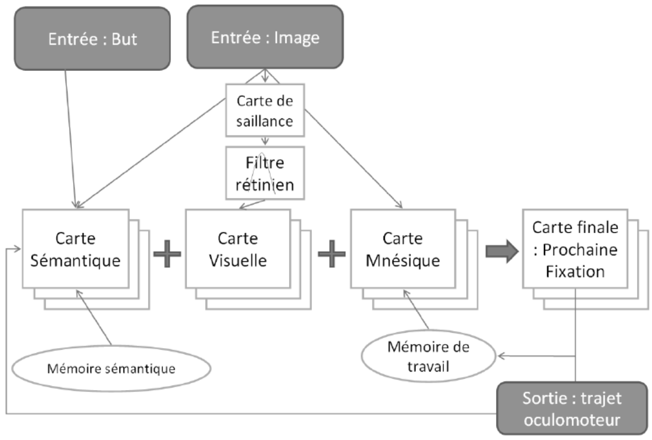


Figure 2.10: Architecture of the computational cognitive model of information search (Chanceaux, 2009).

way the semantic association of the words actually processed in the text and the search goal would be treated as the participant's perception about the relatedness of the text processed so far. In this manner our expectation would be to get a still better model performance by considering a semantic component among others. In the current thesis, we are planning to include the semantic association between words and the search goal as a key component of the computational cognitive model.

2.2.2.4 A computational cognitive model of information search

Chanceaux (2009) presents the development of a computational cognitive model of eye movements of people during a task of information search. She studied the combination of visual, semantic and memory processes in information seeking task on textual interfaces such as web pages. She presented a computer program based on theoretical cognitive models replicating human behavior. She modelled the cognitive processes involved in a task of information search and tried to replicate the participant eye movements recorded during different experiments. Basically, this PhD thesis proposed a model composed of visual, semantic and memory processes. Figure 2.10 shows the general architecture of the model. The visual component is guided by saliency maps which predict areas in the display attracting attention, according to low level information (color, orientation and contrast), and the physiological properties of human vision. This visual map includes two different processes: the visual saliency and the visual module. The visual saliency is presented as in Itti & Koch (2000) where items which are more salient tend to attract fixation (saliency map). The visual module corresponds to the visual acuity (retinal filter), used to avoid long-range saccades moving from one side of the display to the other one, which is not how humans behave. Combining these two mechanisms each item receives a visual weight. On the semantic component (semantic map), the similarity is computed by the model using Latent Semantic Analysis (LSA) (Landauer et al., 2007). LSA is a mathematical tool where words occurring in similar contexts are represented by similar vectors. The semantic similarity between two set of words is computed by using the cosine function between the corresponding LSA vectors. The higher the cosine value, the more similar the words (or sentences) are. Finally and before to execute the next fixation, the memory processess (memory map) are generated by the mechanism of inhibition of return and the variable memory model (Chanceaux, 2009). The whole method is composed of the following steps: designing a theoretical model of interaction, designing a simulation tool, and developing psychophysical experiments with eyetracking techniques to validate and to refine the proposed model.

Differently to the work presented so far, our plan is to study in a more detail the way people make decisions to stop reading an item information (a paragraph) when they are looking for a specific information. This is one limitation stated by Chanceaux (2009) of the model presented because it does not consider an accuracy mechanism to stop reading a paragraph. Additionally, we have simplified the stimuli pages (to 1 or 2 paragraphs per page). This situation will allow us to study the intra-paragraph (e.g. how the words forming a paragraph are processed) strategy of a participant with more accuracy than Chanceaux (2009) because we will be able to manage the stimuli pages with a higher spatial resolution among other advantages.

Finally, after having reviewed most of the important aspects of the searching activity on textual material and some of most representative computational models, we are going now to discuss the way in how the information search is interwined with the decision making. Firstly, behavioral aspects of decision making activity are presented and secondly, a computational model for making such decisions is also presented.

2.3 Decision making and information search

2.3.1 Decisions

A decision is a commitment to a proposition among multiple options. Often this commitment leads to a particular action. It might be said that the life of an organism consists of a serie of decisions made in time (Sartre, 1984). Decision making involves accumulation of evidence about choice alternatives. This process often takes time when the quality of information is poor or there are numerous choice options to consider. In everyday life people often make decisions. We can cite some typical examples by imagining that someone is wondering about the following situations:

- *Shall I bring the umbrella today?* The answer may depends on something unknown, because I do not know if it will rain or not;

- *I am looking for a house to buy. Shall I buy this one?.* It could be the case that I have already found one, but perhaps I will find a still better house for the same price if I go on searching. *When shall I stop the search?*

- *I am in a bookstore to look for a new book to read during my spare time and I have pre-selected two or three of them, but I would like to buy only one. Which one seems to be the most interesting? or should I buy the cheapest one?*

The list of examples would be unending but the three examples show us a wide variety of making decision situations in which people are often faced in real life. Particularly the last two can be seen as information search tasks involving mechanisms of decision making. Most of problem solving and decision making behavior relies on information search. In the current thesis, we are interested in this kind of information search tasks that encourage to adopt decision making mechanisms.

Searching information on the Internet has become a significant activity in the daily lives of individuals in all the world. The expectation is that this activity is going to continue increasing in the future. Very little is known about why people stop searching for information online despite the increasing importance of online search behavior in decision making and problem solving (Browne et al., 2007). Understanding why people stop information search is important to both theory and practice in a wide variety of domains as Browne et al. (2007) have stated. Information search has a direct impact on the number and quality of alternatives considered and on the alternative selected. Stopping a search on the World Wide Web is even more critical due to the high number of existing alternatives. Thus, understanding the mechanisms people use to stop online search is of particular importance. Web designers will be also benefit from getting a better understanding of consumer stopping behavior and they will be able to develop features that will cause people to stop on their sites. In a problem-solving or decision-making process, the search process is terminated at some point because the person judges that she has enough information to move to the next stage (Browne et al., 2007). This judgment of sufficiency is made by using a stopping rule or heuristic (Browne et al., 2007). The stopping rules can be either cognitive or motivational in origin. Cognitive stopping rules result from the ways in which people process information and their mental models of tasks or environments. Preferences, desires, or internal/external incentives, such as deadlines, costs, or preferences for closure give as result the motivational rules. Here, we are only interested in the cognitive stopping rules.

Cognitive stopping rules are used during different stages of decision making. In the early stages dominated by information search and the design of alternatives, they are used to terminate the information acquisition process and to assess the sufficiency of the information collected.

In the latter stages of decision making dominated by choice, stopping rules are used by the decision maker to stop his evaluation of alternatives and to make a choice (Simon, 1996). Four cognitive stopping rules investigating sufficiency of information were first suggested by Nickles et al. (1995) and a fifth was suggested in exploratory data by Browne & Pitts (2007) and verified through empirical investigation by Browne et al. (2005). The five rules are described and exemplified in Table 2.1.

In their study, Browne et al. (2007) used as a working hypothesis the fact that the degree of task structure and the cognitive strategy the decision maker uses to represent the task would be useful in helping to understand the stopping rule use in web-based search. Task structure refers to the degree to which the necessary inputs, operations on those inputs, and outputs are known and recognizable to the decision maker. This is the case of the example shown in Table 2.1 (on first row) where people has a mental list of items that must be satisfied before he will stop collecting information. With this well-structured task, decision makers understand the nature of the problem and the steps necessary to solve it. Here, a descomposition strategy could be used by people because the information available is at least in part discrete and various task elements, criteria, or attributes can be separately identified.

However, people sometimes search for information until their mental representation stops shifting and stabilizes as in the case of the example in Table 2.1 (second row), when the task is poorly structured and it is represented as a whole. Here, the information available may or may not be discrete but the mental representation is integrative. People act based on their sense or image of the situation rather than on individual elements. This kind of strategy is also called holistic.

Rule	**Description**	**Example**
Mental list	Person has a mental list of items that must be satisfied before he will stop collecting information.	In searching for information concerning the purchase of a car, a person continues until he satisfies all the elements on his mental list (e.g., model, price, and color).
Representational Stability	Person searches for information until his mental representation stops shifting and stabilizes.	To diagnose a patient's illness, a physician asks the patient to describe his symptoms. When the physician reasons that his mental model of the patient's condition is no longer changing, he stops asking the patient further questions.
Difference Threshold	When person stops learning new information, he stops his information search (there is a priori difference level to establish that nothing new is learning).	In gathering requirements for a new information system, a systems analyst interviews users until he determines that he is no longer learning new information. At that point, he terminates the requirements elicitation process.

Magnitude Threshold	Person has a cumulative amount of information that he needs before he will stop searching. The focus is on having "enough" information.	When perusing a newspaper article, a reader may skim the article until he has enough information to develop an idea of what it is about.
Single Criterion	Person decides to search for information related to a single criterion and stops when he has enough information about that criterion.	To choose a university to attend, a high school student searches only for information concerning the quality of the universities' overseas programs.

Table 2.1: Cognitive stopping rules adapted from Browne et al. (2007).

The strategy a person will follow depends on several factors such as the task complexity and experience with the task. Browne et al. (2007) operationalize the complexity of the task by using a typology proposed by Campbell (1988). They have also considered two extra hypotheses: tasks with lower complexity in which the decision maker has experience will tend to be approached using a decomposition strategy and tasks with higher complexity in which the decision maker has less experience will tend to be approached using a holistic strategy.

For their study purposes, three different experiments of information search online were carried out. In the first experiment, they asked participants to search for a particular product at BestBuy.com. The task was well-structured, since the inputs, operations, and outputs of the problem should have been recognized and easily understood by participants. There were various alternatives to be evaluated according to a set of decomposable criteria such as price and product features. This was a low complexity task according to Campbell's typology (1988).

In a second experiment, participants were asked to search for a job with Amazon.com. The task was also well-structured because participants should clearly understand the nature of the problem and what is required of them. Participants should also have experience with this type of task. The task was of medium complexity according to Campbell's typology (1988), since there are multiple paths to a desired outcome. The task also contained decomposable elements such as location, salary, job responsibilities, etc.

For these two experiments, their expectation was to have more participants using the mental list and single criterion stopping rules than others because they are well-structured and decomposable tasks.

In the last experiment, they chose a search problem that was poorly structured and, due to its complexity and participants' lack of experience with such tasks, not easily decomposable. They asked participants to search online for the map of a battlefield. They referred to a relatively small battle area, with scarce maps online and varied. Participants were also told that they would need to draw the battlefield from memory after performing their search. The expectation for the experiment was to have more participants using the magnitude threshold and representational stability stopping rules than others because it was a poorly structured and a holistic task.

The experimental results confirmed their hypotheses and suggested that the dimensions of task structure and nature of the person's representation has an important influence on the stopping rules used by people to gauge the sufficiency of information in such tasks.

However, they have studied only two types of search tasks using only two combinations of the important dimensions for determining the stopping rule used: well-structured/decomposable and poorly structured/holistic. As the same authors recognise, further studies with other possible combinations are needed.

They also investigated only the task characteristics of structuredness and representation strategy. Maybe other task characteristics may also has an impact over stopping rules. Another important thing is that the task structure and representation strategy were treated as discrete categories but they are actually continua. Last but not least, new exploratory studies with different stimuli might reveal the use of additional cognitive stopping rules.

2.3.2 Models for decision making

Typically, people need to perform a fast "on-line" text document *classification* for large numbers of documents. For example, we can suppose that we need to do a first *classification* of the most representative research works done in operating systems area published on the last two years. Clearly, it is an interesting problem for cognitive science, because it involves real world human decision making with complicated stimuli. In only two years the number of research work reported could be huge. This is also an important problem in machine learning and related areas which involves deciding whether or not a document is about a given topic. There are automatic classifiers but most of them takes time to process large text corpora (e.g. on documents with a large number of words) because they consider every word in the treated document. In this section, a computational modeling work about the text classification problem is presented with a potential application in an automatic system of text classification.

Lee & Corlett (2003) present an interesting work which involves the central problem of text classification: deciding whether or not a document is about a given topic. They developed two cognitive models of human decision making, and also used them to test automated text classification systems. The following observations are considered:

- People often make non-compensatory decisions because they do not consider all of the words in a document to decide whether or not the text is about a topic.
- The relationship between the decisions "the document is about the topic" and "the document is not about the topic" is a choice between two competing models. The content of the document is used as evidence in favor of either decision.
- People generate more information than just a binary choice when deciding whether or not a text document is about a topic because there is a complete decision-making. They take a period of time for answering and can express a level of confidence in such decision. These measures may be useful for the process of model development, evaluation and comparison.

Lee & Corlett (2003) used as evidence measure, how often a word occurs in documents about a topic relative to how often it occurs in documents not about that topic. For example, the presence of a word like "vatican" in a document provides strong evidence that the document is about the the ecclesiastical or sacerdotal-monarchical state, because it occurs in documents about the topic, and rarely in documents that are not about the topic. In an opposite way, a word like "saccade" provides strong evidence against a document being about the sacerdotal-monarchical state, because it seldom occurs in documents about the topic, but does appear in documents about other topics. Very short words like "of" or "the" provides little evidence in favor of either decision, because they occur at the same rate in documents both about and not about the topic. Authors used the standard Reuters-21578 text corpus (Lewis, 1997) as source of real world text documents.

The models are based on random walk and accumulator sequential sampling processes. In the first one, the total evidence is calculated as the difference between the evidence for the two competing alternatives. A decision is made once it reaches an upper or lower threshold. In this model a measure of confidence in the decision is determined by the number of words examined. The higher the number of words required to classify a document, the lower the confidence and viceversa. The overall modeling process is interpreted in Bayesian terms and the final formulation considers both prior probabilities of "yes" and "no" decisions that determine the initial point of the random walk, followed by the summation of the evidence provided by each successive word in the document. Figure 2.11 summarizes an example of the operation of the random walk model on a document that is about the topic being examined. The state of the random walk is shown as the evidence provided by successive words in the document. A threshold value of 50 is shown by the dotted lines above and below. Particularly, the example shows very well the use of non-compensatory decision making because if only the first 100-260 words of the document have been treated, a correct "yes" decision would be

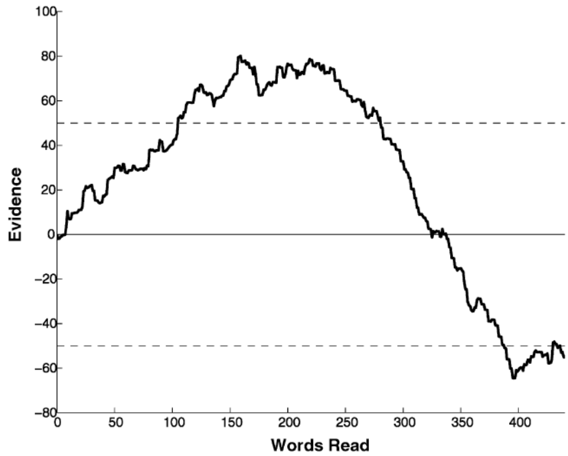


Figure 2.11: Example of the operation for the random walk model when the document is about the topic (Lee and Corlett, 2003).

made because the evidence value is upper than the threshold but the final state of the random walk when the entire document ($\sim$ 450 words) has been considered, favors a "no" decision being made because the evidence value is lower than the threshold.

Differently, in the accumulator model both separate evidence totals for the "yes" and "no" decisions are computed. As with the random walk model, these totals may begin at non-zero values to reflect decision bias, and then accumulate evidence by reading the words in the document. When the i_{th} word is read, the two evidence totals are updated. More specifically, the evidence provided by each successive word is added to the "yes" accumulator if it is positive or the "no" accumulator if it is negative. Once either the "yes" accumulator or the "no" accumulator reaches certain threshold value, the corresponding decision is made. Confidence may be also determined by the number of words read. However, a good alternative of confidence measure may be assessed using a "balance of evidence" approach (Zandt et al., 2000), where it is measured as the difference between the evidence totals as a proportion of the total evidence accumulated. Documents classified at the same point could be given different confidence measures depending on the level of evidence in the "no" accumulator: if the "no" accumulator contains no evidence, confidence will be high, whereas, if it contains almost as much evidence as the "yes" accumulator, confidence will be low.

Fig. 2.12 shows the operation of the accumulator model on the same text document considered in Fig. 2.11. Both accumulator totals are shown (by the two curves) as successive words in the document are read, and thresholds of 50 are once again indicated by dotted lines. As with the random walk model, the accumulator model makes a non-compensatory "yes" decision, because the "yes" accumulator is the first to reach its threshold (at word 90^{th}). After all of the words in the document have been read ($\sim$450

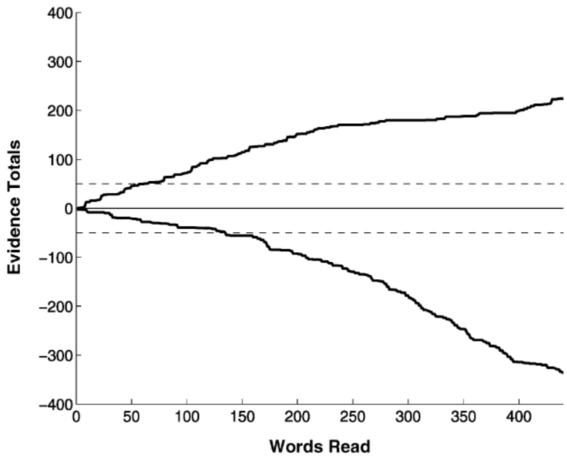


Figure 2.12: Example of the operation for the accumulator model on the same document shown in Fig. 2.11 (Lee and Corlett, 2003).

words), however, a "no" decision is favored, because the "no" accumulator has greater absolute value of evidence (∼350).

The two models were created and validated with experimental data recorded in a simple experiment. In the experiment, a limited form of human text classification is considered, where words are presented serially (word by word) at a constant rate (one word per second) until a decision is made regarding whether or not the document is about a given topic. Two different instructions were given to the participant to make the decision, either as quickly as possible or with the best possible accuracy. After each decision, participants were required to express their confidence in such decision.

Experimental results showed what they expected which is that people's text classification decisions are non-compensatory because their decision making is not done by reading all the document words. Saying differently, their decision making is not necessarily subject to a speed-accuracy tradeoff.

The random walk and accumulator model parameters were fitted to the empirical data: the evidence thresholds for the "yes" and "no" decisions. Fitting work is done for both models for the speed and accuracy conditions. For the accumulator model, a large number of parameters matched all of the participants' decisions for both the speed and accuracy conditions. However, for none of the parameterizations examined did the random walk model correctly predict the participants' decisions for either the speed or accuracy data.

They explain this by focusing on an important difference between the two models. The accumulator could make a "yes" decision, because the corresponding accumulator is the first to reach the threshold, whereas the random walk model never reaches the positive threshold, and eventually makes a forced "no" decision when all of the words in the document have been read. The situation is because random walks treat evidence in favor of alternative decisions as being more commensurable than accumulators, in the sense that the presence of evidence in favor of one decision can be directly negated by equal strength evidence in favor of its alternative. For that, the evidence computation process used by the accumulator seems better suited than the random walk to modeling the decisions made by humans in classifying text documents. It is possible, of course, that more sophisticated sequential sampling models based on random walks could fit better the decision data.

The performance of the accumulator model presented is promising. However, the evidence value used is a very simple model of human semantic representation. When people read text documents, meaning can be learned from parts of words, or sequences of words, in a way that is mediated by the earlier content of the document. Providing sequential sampling models with evidence values that captured some of this richness in meaning seems likely to improve their performance and this is not exactly the case of the evidence value described in this work. Some examples of such models are Latent Semantic Analysis (Landauer et al., 2007) and Hyperspace Analogue to Language (Lund and Burgess, 1996) approaches, as well as the probabilistic method developed by Griffiths & Steyvers (2002), all of which measure, in various ways, the patterns with which words appear in the same and different contexts.

Moreover, both the random walk and accumulator models presented here do not involve any form of memory, nor do they adapt to their environment in any self-regulating way and, clearly real world human text classification involves a limited memory and a potential to learn and adapt. The extension of the models to incorporate these characteristics is theoretically important.

An important difference with respect to the aim of the current thesis is that we are interested in studying people when they are engaged in task of information search and by presenting the whole stimuli at a time (a full article or paragraph). Let us saying that the cognitive behavior of people should be observed in a more ecological manner and not by presenting one word at a time as it was done by Lee & Corlett (2003).

2.4 Statement of the problem

Our research problem is placed into the computational cognitive modeling of people's behavior faced to an information search task. The aim is to get a better understanding of the cognitive processes involved during a task of information search on textual material by using a cognitive computational model-based approach. With such a model we will be able to make predictions about how a textual document would be processed when people are looking for information. There are also some practical benefits from such a cognitive model because knowing where and why a user is likely to gaze at during an information search task is important for the design and evaluation of complex documents and especially web pages. Many works have already been done in this area, in particular in the context of web usability, but they usually consider that the user is processing the page without any specific goal or by presenting the stimuli in a very artificial manner, among others restrictions, as we have presented in this chapter.

In the current thesis, we are interested in a modeling work for a very specific information search task which always involves a goal (eg. a theme, a word, etc.) and concerned with deciding whether an item of information read (eg. a paragraph, a newspaper article, etc.) is related or not to that search goal or whether the current item is more interesting or not than another item that has been processed previously. Since the idea is to observe people's behavior in the most possible ecological way like they were faced to a real task, psychophysical experiments with minimum restrictions (eg. a self-paced manner reading) are designed in order to study, create and validate the model. Two important tools currently used in cognitive science research were considered in the design of our experiments: eye-tracking and EEG techniques. But the development of our computational cognitive models is only based on eye-tracking data.

Chapter 3
Method

In the earlier chapter, we have stated the main goal of the current thesis: the development of two computational cognitive models to simulate people's behavior when they are faced to two different information search tasks. More specifically, in both cases we are interested in modeling the particular decision making to stop reading an item of information. Information search can be made on any kind of documents, but here the interest is mainly on textual documents or paragraphs. As we have already pointed out previously, information search is different from pure reading because people have a goal in mind while processing a document. They have to constantly keep in memory this additional information.

Since we followed an experimental-based approach, we carried out psychophysical experiments in order to acquire data on eye movements, create and validate the models. The idea of the current chapter is therefore to explain in a general way the methodology that we followed to achieve our goals. We start by discussing the conception of the experiments and then present the construction and validation of the models. A comprehensible explanation of the tools that were used along the overall process is also given. Now, we start with a general explanation of the designed experiments.

3.1 Experiments

We designed two different experiments, each one for a different situation of information search.

Situation 1 (one goal, one paragraph): People are concerned with this situation in everyday life. For instance, you are looking in a cookbook for a Mexican recipe. Here the recipe you are currently reading and the goal (a Mexican recipe) are the pieces of information involved and both have to be together managed in order to make a correct decision about if the recipe match well your search criteria. During our experimental work, paragraphs were the items or blocks of information considered. Then participants were asked to read a paragraph with the task of quickly deciding if it is related or not to a given goal. Figure 3.1 illustrates the situation we aim at modeling.

Situation 2 (one goal, two paragraphs): People are also typically concerned with this situation. For example, you are looking in a cookbook for a nice French recipe and

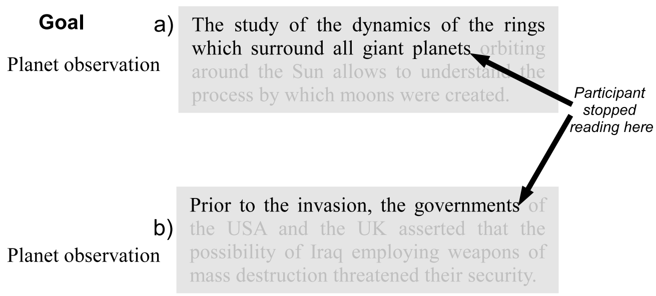


Figure 3.1: Illustration of the two input data of the model for SITUATION 1: the goal (*Planet observation*) and the paragraph. The paragraph is abandoned before its end because enough information has been gathered and maybe due to a) a high-relatedness to the goal b) a low-relatedness to the goal.

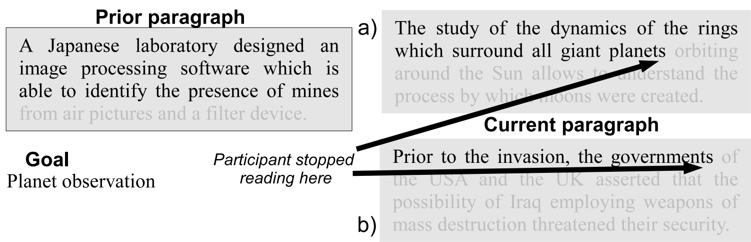


Figure 3.2: Illustration of the three input data of the model for SITUATION 2: the prior paragraph, the goal and the current paragraph. The current paragraph is abandoned before its end because enough information has been gathered and maybe due to a) a high-relatedness to the goal and enough information about the difference between current and previous paragraphs. b) a low-relatedness to the goal.

you already found one but you want to find a better one. Here things are different because at least three pieces of information have to be together managed in order to make a correct decision: the current recipe, the goal (a nice French recipe) and a previous recipe (already found). During our experimental work, paragraphs were also the items of information considered. Then, participants were asked to read a paragraph with the task of quickly deciding whether it is better related to a given goal than another paragraph processed previously. Figure 3.2 illustrates the situation we aim at modeling.

In both situations we aim at focusing on a behavior that is specific to information search, which is stopping processing a paragraph before it is completely read. We suspect that people should decide to stop reading a paragraph in a different manner. For example, in SITUATION 1, they may decide to stop reading a paragraph when they consider that they have collected enough information about the paragraph. This can be either due to a high-relatedness or a low-relatedness to the goal as is shown in Fig. 3.1. In Situation 2, things are different because people are involved in a comparative task and they may decide to stop reading a paragraph when their perception about the difference between both paragraphs (the previous and the current) is large enough to decide which one is more related to the goal. In a similar way, the decision to stop reading the current paragraph might be also motivated because of its high-relatedness or its low-relatedness to the goal as is shown in Fig. 3.2

One experiment was carried out for each of the situations. The Experiment 1 involved both eye movement and brain signal (EEG) recording because it is a part of the research project ANR Gaze-EEG "Joint synchronous EEG signal and eye tracking processing for spatio-temporal analysis and modeling of neural activities". However, and for our modeling purposes only the eye movement data were analysed. In this experiment, our interest was to study how people make a binary decision to keep or reject a paragraph according to its relatedness or not with a search goal. We made the consideration of having an experiment with a very low variability (eg. with all participants exposed to the same stimuli pages, using simple stimuli pages, etc.) as a desirable requirement because of the joint analysis of eye-tracking and EEG data. For that reason, in the experiment, the sequence of stimuli pages presented is independent of the participant's answer and it was only managed by a random generator.

The Experiment 2 only involved the eye-tracking technique. The idea behind the Experiment 2 was only to study people's behavior when they are faced to a comparative task of deciding if the current paragraph is more interesting than another paragraph processed previously. Here, the complete sequence of stimuli pages presented on the screen depended of the participant's answer and of a random choice during trials. In other words, participants traced their *own route* along the experiment.

For the construction of the material used for the experiments a model of human semantic judgements was required. The relatedness of a paragraph to the goal was controlled by using Latent Semantic Analysis (Landauer et al., 2007) or simply LSA. We would like to start by explaining this model before giving a description of the experimental material.

3.2 Latent Semantic Analysis (LSA)

LSA is a fully automatic mathematical/statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. It is not a traditional natural language processing or artificial intelligence program (Landauer et al., 2007).

In a first stage LSA takes a large corpus of text as input (eg. all articles published in a newspaper during one year) and generates an occurrence matrix F of m words x n documents. Each nonzero element f_{ij}, of the matrix F is the occurrence frequency

Label	Document
M1	*Rock* and *Roll Music* in the 1960's
M2	Different *Drum Rolls*, a Demonstration of Techniques
M3	*Drum* and Bass *Composition*
M4	A Perspective of *Rock Music* in the 90's
M5	*Music* and *Composition* of Popular Bands
C1	How to Make *Bread* and *Rolls*, a *Demonstration*
C2	*Ingredients* for Crescent *Rolls*
C3	A *Recipe* for Sour*dough Bread*
C4	A Quick *Recipe* for Pizza *Dough* using Organic *Ingredients*

Table 3.1: Example of documents about music (M) and cooking (C).

Word/Document	M1	M2	M3	M4	M5	C1	C2	C3	C4
Bread	0	0	0	0	0	1	0	1	0
Composition	0	0	1	0	1	0	0	0	0
Demonstration	0	1	0	0	0	1	0	0	0
Dough	0	0	0	0	0	0	0	1	1
Drum	0	1	1	0	0	0	0	0	0
Ingredients	0	0	0	0	0	0	1	0	1
Music	1	0	0	1	1	0	0	0	0
Recipe	0	0	0	0	0	0	0	1	1
Rock	1	0	0	1	0	0	0	0	0
Roll	1	1	0	0	0	1	1	0	0

Table 3.2: Occurrence matrix corresponding to the example of the Table 3.1

of ith word in the jth document. Table 3.1 shows a simplified example (Martin and Berry, 2007) of documents and its occurrence matrix is shown in Table 3.2. Then, the matrix is normalized in order to make emerge the most meaningful words by applying local and global weighting functions:

$$f\prime_{ij} = local_{ij} \times global_i$$

A typical transformation used for this purpose is the *log*-entropy function which represents a *log* transformation at local level and a entropy transformation at global level. Then the local and global functions are defined as:

$$local_{ij} = log(f_{ij} + 1)$$

$$global_i = 1 + \frac{\sum_j p_{ij} log(p_{ij})}{log(n)}$$

with

$$p_{ij} = \frac{f_{ij}}{\sum_j f_{ij}}$$

In these equations f_{ij} corresponds to the number of occurrences of the i word in the document j, and n is the total number of documents.

Once the matrix was normalized, the last stage in the transformation of the matrix is the key point in the method. Here, the matrix is reduced in order to make emerge the relationships between words although they are not co-occurring. One of the most classical approach used is the *Singular Value Decomposition* (SVD) of matrix A. By this way, the matrix is decomposed into a product of three matrices:

$$A = U \Sigma V^T$$

with U and V are the orthogonal matrices and Σ is a diagonal matrix. Fig. 3.3 shows a schematic representation of a SVD transformation. The matrix U corresponds to the words and the matrix V to the documents. The matrix Σ contains the singular values, sorted in descending order. The resulting matrices of our example are showed in Tables 3.3, 3.4 and 3.5. In order to remove noise, the matrix is reduced to a fewer number of dimensions. This is done by canceling the coefficients of the diagonal matrix, starting with the smallest one. In our example and in order to have a two-dimensional representation, only the two first singular values (the biggest ones) are kept (they are bold-faced in Table 3.4) and give us the graphical representation showed in Fig. 3.4. However, it is worth noting that the two-dimensional representation is only for illustrative purpose since that it would not work with a real French or English corpus. In the figure 3.4, words that are semantically close are also spatially close to each other, even if they do not co-occur in the same document as *Rock* and *Demonstration*. In the same way, it is possible to observe the similarity in between documents.

The number of dimensions (in general $k \approx 300$ for English or French) was determined in a empirical way by different tests. A very small number of dimensions results in an information loss and a very high number of dimensions does not allow to make emerge the semantic relationships between the words. With the resulting matrices, we are able to use LSA and this procedure is now explained.

3.2.1 How the LSA model is used?

The semantic similarities between two set of words can be easily computed by using the matrices already described since we have a two-dimensional vector representation for each word. For instance, vector representation of word *Roll* in our tiny example is

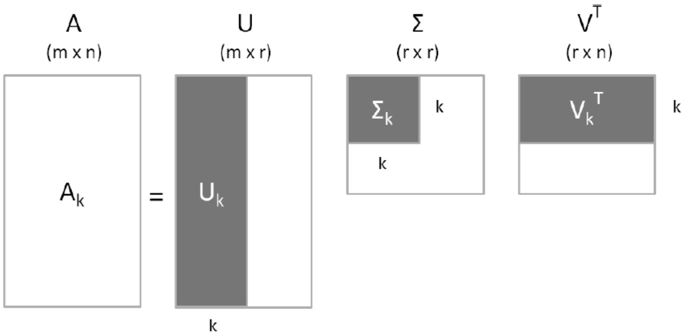


Figure 3.3: Schematic representation of a SVD transformation.

Bread	**.42**	**-.09**	-.20	.33	-.48	-.33	.46	-.21	-.28
Composition	**.04**	**-.34**	.09	-.67	-.28	-.43	.02	-.06	.40
Demonstration	**.21**	**-.44**	-.42	.29	.09	-.02	-.60	-.29	.21
Dough	**.55**	**.22**	.10	-.11	-.12	.23	-.15	.15	.11
Drum	**.10**	**-.46**	-.29	-.41	.11	.55	.26	-.02	-.37
Ingredients	**.35**	**.12**	.13	-.17	.72	-.35	.10	-.37	-.17
Music	**.04**	**-.35**	.54	.03	-.12	-.16	-.41	.18	-.58
Recipe	**.55**	**.22**	.10	-.11	-.12	.23	-.15	.15	.11
Rock	**.05**	**-.33**	.60	.29	.02	.33	.28	-.35	.37
Roll	**.17**	**-.35**	-.05	.24	.33	-.19	.25	.73	.22

Table 3.3: Matrix U: Word vectors

1.1	0	0	0	0	0	0	0	0
0	**.96**	0	0	0	0		0	0
0	0	.86	0	0	0	0	0	0
0	0	0	.76	0	0	0	0	0
0	0	0	0	.66	0	0	0	0
0	0	0	0	0	.47	0	0	0
0	0	0	0	0	0	.27	0	0
0	0	0	0	0	0	0	.17	0
0	0	0	0	0	0	0	0	.07
0	0	0	0	0	0	0	0	0

Table 3.4: Matrix $\sum$: Singular values

M1	**.07**	**-.38**	.53	.27	.08	.12	.20	.50	.42
M2	**.17**	**-.54**	-.41	.00	.28	.43	-.34	.22	-.28
M3	**.06**	**-.40**	-.11	-.67	-.12	.12	.49	-.23	.23
M4	**.03**	**-.29**	.55	.19	-.05	.22	-.04	-.62	-.37
M5	**.03**	**-.29**	.27	-.40	-.27	-.55	-.48	.21	-.17
K1	**.31**	**-.36**	-.36	.46	-.15	-.45	.00	-.32	.31
K2	**.19**	**-.04**	.06	-.02	.65	-.45	.41	.07	-.40
K3	**.66**	**.17**	.00	.06	-.51	.12	.76	.25	-.35
K4	**.63**	**.27**	.18	-.24	.35	.10	-.35	-.20	.37

Table 3.5: Matrix V: Document vectors

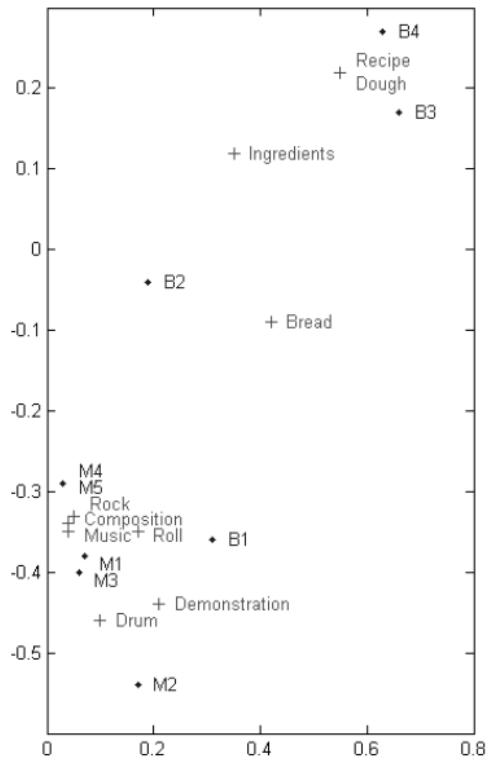


Figure 3.4: 2-dimensional spatial representation of data after the SVD transformation.

$(.17, -.35)$. Moreover, such a vector formalism is very convenient to give a representation to sentences that are not in the corpus: the meaning of a new sentence is represented as a linear combination of its word vectors. For example the vector representation of the sentence "*Recipe Composition*" will be: $(.55, .22) + (.04, -.34) = (0.59, -0.12)$. Therefore, any sequence of words can be given a representation. The semantic similarity between two sequences of words can be computed using the cosine function of their angle:

$$similarity = cos(\theta) = \frac{Words_1 \cdot Words_2}{||Words_1||||Words_2||}$$

with $Words_1$ and $Words_2$ are the LSA word vectors corresponding to the two set of words. The higher the cosine value, the more semantically similar the two sequences of words. A cosine value of 1 indicates a maximum similarity whereas a value of 0 (or a very small negative number) indicates a null similarity.

We can make some similarities computation by using our example. We have the following word vectors in the 2D space: *Music*= $(.04, -.35)$, *Roll*= $(.17, -.35)$, *Drum*= $(.10, -.46)$, *Demonstration*= $(.21, -.44)$, *Recipe*= $(.55, .22)$ and *Bread*= $(.42, -.09)$. Now, by computing the cosine between the LSA vectors corresponding to *Music* and *Roll*, we got a value of 0.94. By doing the same for the words *Recipe* and *Bread*, we got a value of 0.83. The semantic similarity between *Music* and *Roll* is higher than the corresponding cosine between *Recipe* and *Bread*. These values are expected because from Fig. 3.4 the words *Music* and *Roll* are spatially closer than the words *Recipe* and *Bread*. In a second example, we can compute the semantic similarity between two set of words: {*Music*,*Roll*}, {*Recipe*,*Bread*} and {*Drum*,*Demonstration*} by adding the word vectors corresponding to each word as follows: A = {*Music*,*Roll*}= $(.04, -.35) + (.17, -.35) = (0.21, -0.70)$, B = {*Recipe*,*Bread*}= $(.55, .22) + (.42, -.09) = (0.97, 0.13)$ and C= {*Drum*,*Demonstration*}= $(.10, -.46) + (.21, -.44) = (0.31, -0.90)$. Then, the semantic similarity between *Music-Roll* and *Recipe-Bread* is 0.16 and in between *Music-Roll* and *Drum-Demonstration* is 0.99. Values are also consistent because *Music-Roll* and *Drum-Demonstration* are spatially much closer in between (all of them are semantically related to the *Music*) than *Music-Roll* and *Recipe-Bread*.

We have explained here the way LSA method works by using a very simplified example and considering only two dimensional LSA vectors. However, for our experimental and modeling purposes LSA was trained on a 24 million word French corpus composed of all articles published in the newspaper "Le Monde" in 1999 and we used 300 dimension LSA vectors as usual.

Once explained how the LSA model works, the description of material used during the experiments is presented.

3.3 Material

We kept the material as it was used in an older experiment carried out in our research team. They were created a set of 30 goals (topics). Each one was expressed by a few words (e.g. *mountain tourism*, *planet observation*, etc.). For each goal, 7 paragraphs were created (with mean=30.1 words, σ=2.9) with all the text in French. Since that the idea was to study the way in that people judge the relatedness or not of a paragraph with a given goal, we ended up with two paragraphs highly related (HR), two paragraphs moderately related (MR) and three paragraphs unrelated (UR) to the goal. LSA (Landauer et al., 2007) was used to control the relatedness of a paragraph to the goal. For the HR texts, cosine with the goal was above 0.40, for MR texts, cosine was between 0.15 and 0.30, and for UR texts, cosine was below 0.10. The three paragraphs UR were used during our Experiment 2 (the first experiment carried out) but only two of them were used for our Experiment 1. Paragraphs HR and UR represent the extreme cases. Participants should easily decide that a paragraph HR is related to the goal and that a paragraph UR is not related to the goal. Paragraphs MR are in between, and they were considered because of their role of neutrality and difficulty to discern about their relatedness with the corresponding search goal. People behavior in these cases would be unpredictable. The stimuli pages were generated with a software that stored the precise coordinates of each word on the screen.

Figure 3.5 shows the English translation of three texts (HR, MR, and UR) for the goal "observation des planètes" (planet observation).

The study of the dynamics of the rings which surround all giant planets orbiting around the Sun allows astronomers to understand the process by which solar system moons were created.

(a) A highly related text (HR).

A Japanese laboratory team designed an image processing software which is able to identify the presence of mines from air pictures and a filter device.

(b) A moderately related text (MR).

The complaint was filed by the company and concerned the street furniture and the self-service bike renting system. The American group lodged an appeal.

(c) An unrelated text (UR).

Figure 3.5: Examples of the textual material used for the experiments (translated into English). The goal is "planet observation".

For our measurement purposes, the description of the apparatus used during the experiments is now presented.

3.4 Apparatus

As mentioned previously, during our Experiment 1, a synchronized recording of both eye movement and EEG signals was done but in our modeling work we are only concerned with the eye movement data. Eye movements were recorded by using a remote binocular infrared eyetracker EyeLink 1000 (SR Research). The EyeLink system was used in the Pupil-Corneal Reflection tracking mode sampling at 1000Hz. In Experiment 2, only the eye movements were recorded by using a head mounted eye tracker EyeLink II (SR Research), sampling pupil position at 500 Hz. We wrote our experiments in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Due to the fact that the eye-tracker generates a platform-portable binary record of eye-position and synchronization events, namely a EDF file (Eyelink Data File), we had to perform a pre-processing stage in order to adequate this output to a more accessible format for our further analysis. This pre-processing stage is now explained.

3.5 Pre-processing data

From eye-tracker data to a fixation file. The EDF file may be translated by the EDF2ASC utility (EDF to ASCII converter) into a text-format ASC file. This file lists most of the important data in the EDF file in a more easily accessible format, but at the expense of much larger file size. Below, an excerpt of an ASC file is shown. Numbers at left margin are only for explanation purposes and they are not part of the EDF file.

```
1  ** CONVERTED FROM C:\ExperimentParagraph\Data\s1.edf using edfapi 3.1
      Win32 Mar  2 2011 on Sat Jan 07 14:08:15 2012
2  ** DATE: Fri Jan  6 19:56:28 2012
3  ** TYPE: EDF_FILE BINARY EVENT SAMPLE TAGGED
4  ** VERSION: EYELINK II 1
5  ** SOURCE: EYELINK CL
6  ** EYELINK II CL v4.56 Aug 18 2010
7  ** CAMERA: EyeLink CL Board=0.3 Sensor=BJE
8  ** SERIAL NUMBER: CL1-6CA13
9  ** CAMERA_CONFIG: 6CA13140.scd
10 ** Recorded by EyelinkToolbox demo-experiment
11 **
12
13 MSG 5900361 MATBUILDER VERSION : 2011-08-01
14 INPUT 5900366 0
15 MSG 5900383 DISPLAY_COORDS 0 0 1023 767
16 MSG 5900387 FRAMERATE 60 Hz.
17 MSG 5900387 SUBJECT NAME : s1
18 MSG 5900387 SUBJECT SESSION : 1
19 MSG 5900387 SCREEN WIDTH : 1024
20 MSG 5900388 SCREEN HEIGHT : 768
21 MSG 5900388 SCREEN PHYS DIMX MM : 520
```

```
MSG 5900388 SCREEN PHYS DIMY MM : 320
MSG 5900388 SCREEN DISTANCE MM : 680
MSG 5900389 GUIDING EYE : RIGHT
MSG 5900389 CROSS LENGTH : 15
MSG 5900389 KEEP BUTTON : L
MSG 5900389 STARTING
MSG 5900389 DISPLAY_ON
...
MSG 6138726 START IMAGE : chasse_oiseaux-f1.png
MSG 6138726 VALID FIXATIONS
6138726   240.7   263.0   994.0
6138727   240.8   262.9   996.0
...
6138880   243.7   264.4   991.0
6138881   243.7   263.8   990.0
EFIX R   6138612  6138881  270   241.8   263.0    993
SSACC R  6138882
6138882   243.7   263.3   990.0
6138883   243.8   263.5   990.0
...
6138918   337.6   267.3  1008.0
6138919   337.8   267.6  1009.0
ESACC R  6138882  6138919  38     243.7   263.3   337.8   267.6   3.92   243
SFIX R   6138920
6138920   337.6   267.8  1010.0
6138921   337.4   268.3  1011.0
...
6139039   337.7   273.7   998.0
6139040   337.6   273.9   999.0
EFIX R   6138920  6139040  121   336.5   271.8   1011
SSACC R  6139041
6139041   337.6   274.2  1000.0
6139042   337.5   274.2  1001.0
...
6139063   394.7   275.4   997.0
6139064   394.3   275.8   996.0
ESACC R  6139041  6139064  24     337.6   274.2   394.3   275.8   2.40   203
SFIX R   6139065
6139065  393.1    276.0   996.0
6139066  392.1    276.0   995.0
...
6139267   396.1   275.2   848.0
6139268   396.0   275.0   848.0
EFIX R   6139065  6139268  204   395.2   276.2    913
...
MSG 8410891 DISPLAY_OFF
MSG 8410892 ENDING
```

The ASC file contains general information about the eye-tracker parameters as is shown on line numbers 1-10. From lines 13 to 28, the general information about the subject name, screen dimension, screen resolution and some extra information is also stored. In the experiment where this ASC was recorded, a stimuli image called *chasse_ oiseaux-f1.png* was displayed on the screen as indicated by line number 30. The information stored from line number 32 is what we really need. Basically we have the location of the eyes (or a single eye as in the example) every millisecond (if the eye-tracker frequency is 1000 Hz) or two milliseconds (if the eye-tracker frequency is 500 Hz). These locations are split into saccades and fixations. From line number 32, lines of the file contain the time stamp, the XY location of the eye and the pupil diameter. The line beginning with *EFIX* (End of Fixation) is an indication that the system has recognized the end of a fixation. For example, the line 37 (*EFIX R*) indicates the end of a fixation of the right eye. This is followed by the time of the first (beginning time) and last (ending time) sample in the fixation, the fixation duration, the gaze position data (XY) and the average pupil size. The next line is therefore an indication that there is a beginning of a saccade of the right eye (*SSACC R*: Start of Saccade of the right eye). On line 44, the end of this saccade is indicated (*ESACC R*). Then, a new fixation is started on line number 45 (*SFIX R*) whose end is indicated on line 51 (*EFIX R*). The beginning and ending of a new saccade is indicated on lines 52 (*SSACC R*) and 58 (*ESACC R*) respectively. A new fixation started on line 59 (*SFIX R*) and ended on line 65 (*EFIX R*). The data file its terminated at line number 67-68 with the messages *DISPLAY_ OFF* and *ENDING* respectively.

Since we are only interested in fixations, we have built a program which only extracts fixations from this huge file and generates a text file which looks like this:

```
...
s1 chasse_oiseaux-f1 6138612 6138881 270 241.8 263.0 993
s1 chasse_oiseaux-f1 6138920 6139040 121 336.5 271.8 1011
s1 chasse_oiseaux-f1 6139065 6139268 204 395.2 276.2 913
...
```

The three fixations showed above were extracted from the ASC segment presented early. Each line corresponds to a fixation. Fields are <*subject name*>, <*image name*>, <*beginning time*>, <*ending time*>, <*duration*>, <*X*>, <*Y*>, <*pupil diameter*>. We have another program that takes as input the fixation file, reads the corresponding image file and visualize the experimental scanpath as showed in Fig. 3.6. Fixations are indicated by the small spots whereas the saccades are represented by the lines connecting pairs of fixations.

There is another program which adds to this fixation file the words processed during each fixation. It also adds extra information needed for the further analysis. It takes as input the fixation file plus a file containing the location (XY) of each word in the stimuli-image. The program works with a special mechanism which is now explained.

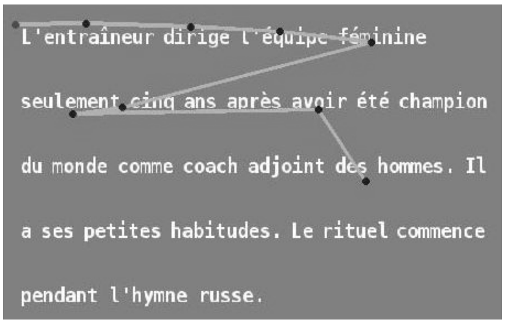


Figure 3.6: Example of a scanpath generated during one of our eye-tracker experiments.

Prediction of words processed during each fixation. In order to predict the words processed during each fixation, we used a window-based approach according to the word identification span (Rayner, 1998) explained in the previous chapter. The size of our window was of 4 characters to the left plus 8 characters to the right of the fixation point. Another important result that we took into account to determine the area from which information is acquired in each fixation is the fact that initial fixations in the beginning part of a word facilitate its recognition more than initial fixations toward the end of the word (Farid and Grainger, 1996). Then, during a fixation, the words catched by the window are candidates to be considered as processed but an extra criteria is applied to decide it: a word is actually processed if at least the first third of it or the last two-thirds are inside the window. Figure 3.7 shows an example of our window based approach to determine which words are processed during a single fixation. The fixation is indicated by the gray circle on the word *des*. This particular example shows that the words *traitement*, *des* and *données* are candidates to be considered as processed. However, using our extra criteria, only the words *des* and *données* would be considered as actually processed. Our program implements this window mechanism by analysing line-by-line the fixation file and by adding the words likely to have been processed by participants during each fixation. Additionally, it also adds more information for a convenience of our modeling purposes. This is the only file used during the modeling work.

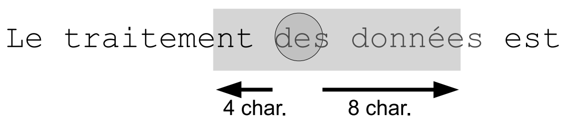


Figure 3.7: Words processed during a single fixation using our window-based approach. The word identification span is limited by the gray rectangle and the fixation point is on the word *des*.

After having presented the way the acquired data are pre-processed in order to adequate them to the modeling work, we would like to explain the approach we followed to build the computational cognitive models considered in the current thesis.

3.6 Model

Two models are developed along the current thesis, one for each of the experiments. However, the approach followed for the construction of both is common and its description is now given.

3.6.1 Statistical approach

In both models the idea was to simulate people's behavior when they are faced with an information search task. We were interested in predicting the moment at which participant would decide to stop reading a paragraph before it is completely read. This may be because they have collected enough information to make a correct decision about the relatedness or unrelatedness of the paragraph with the search goal. Firstly, we started by imagining which variables could play a role in such decision. We believed that the semantic similarity between the words actually processed in a paragraph and the goal must be involved because this variable would be able to mimic human judgments of semantic associations. For that purpose, we used LSA (Landauer and Dutnais, 1997; Landauer et al., 2007) as a model of semantic memory. We also thought that the role of this variable should be different along the two experiments, but here for our general explanation we will refer to it simply as *semantic association* or *Sim*. Naturally, we thought that a high/low value of this variable may induce a decision to stop reading a paragraph because in these cases the perception about the relatedness/unrelatedness of the current paragraph would be clear, but what about if this high/low value appears too early or too late in scanpath? Obviously, the perception would be not the same. We assumed therefore that the decision also depends on the number of fixations made on a paragraph. Even if there is not a one-to-one mapping between fixations and the number of words processed, the number of fixations can be used as a general indicator of the number of words processed in a paragraph. To sum up, the more words processed (more fixations), the higher the confidence in the perception of the relatedness between the paragraph and the goal. Therefore, we assumed that there should be a relationship between *Sim* and a second variable called *rank of fixation* or simply *Rank*.

Once we had determined the two variables that could played a role in the decision to stop reading a paragraph, two empirical fixation distributions were computed: the abandon and the no-abandon distributions. These distributions were computed in the $Sim \times Rank$ space of participant data in order to study how the decision depends on these two variables. To compute the distributions, each participant fixation was associated to a point in the $Sim \times Rank$ space. The abandon distribution is increasing along the *Rank* axis because the probability of abandoning is higher and higher. However, it is the opposite for the no-abandon distribution. A detailed explanation about the computation of both distributions is given in Chapters 4 and 5 for the Experiments 1 and 2 respectively.

The goal was to learn the frontier between both cases in order to be able to predict if a sequence of words already processed is likely to lead to the abandon (stop reading) or the pursuance of the reading task. Depending on the location of any observation (*sim*,*rank*) above or under the curve (frontier), the reader's behavior can be predicted. However, the frontier cannot be just the intersection of the raw data, because of the noise due to the experimental way of gathering data and the obvious drawbacks of the Latent Semantic Analysis method of computing similarities. To find this frontier, a methodology based on a Bayesian inference is therefore used and is presented now.

3.6.2 Bayesian inference

Many aspects of cognition can be formulated as solutions to problems of induction. Given some observation about the real world, we can make conclusions about the process or structure that generated such data and use this knowledge to make predictions for future cases. Bayesian inference is an engine for solving such problems within a probabilistic framework, and consequently is the heart of most probabilistic models of cognition (Griffiths et al., 2008).

The Bayesian inference is based on Bayes' rule. Bayes' rule is no more than an elementary result of probability theory. We can assume the existence of two random variables, A and B. Thankful to the chain rule we can write the joint probability of these two variables taking on particular values a and b, $P(a,b)$, as the product of the *conditional probability* that A will take on value a given B takes on value b, $P(a|b)$, and the marginal probability that B takes on value b, $P(b)$. Thus, we have:

$$P(a,b) = P(a|b)P(b)\,.$$

or we can also write

$$P(b,a) = P(b|a)P(a)\,.$$

From the two equations we can re-write:

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}\,. \tag{3.1}$$

This expression is Bayes' rule, which indicates how we can compute the conditional probability of b given a from the conditional probability of a given b.

Now, in our particular case, let us consider a classification problem with two classes: Abandon (Ab) and No-abandon ($\overline{Ab}$) a paragraph. What is the class $p_{SR}(s, r)$ of a two-dimensional observation (s,r) in the $Sim \times Rank$ space when the posterior probabilities $P(\overline{Ab}|s, r)$ and $P(Ab|s, r)$ are given? The decision rule is:

$$P(\overline{Ab}|s, r) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab|s, r), \tag{3.2}$$

and by using the form of Equation 3.1, we can express both cases as follows:

$$P(\overline{Ab}|s, r) = \frac{P(\overline{Ab}) \times p_{SR}(s, r|\overline{Ab})}{p_{SR}(s, r)} \tag{3.3}$$

and

$$P(Ab|s, r) = \frac{P(Ab) \times p_{SR}(s, r|Ab)}{p_{SR}(s, r)}. \tag{3.4}$$

Then data were regularized by following a statistical parametric approach since they were obviously affected by the noise inherent to acquisition and pre-processing. The statistical model to estimate both density functions for the no-abandon and the abandon cases and the prior probabilities are explained in Chapter 4 and 5 in order to use the Bayesian classifier:

$$P(\overline{Ab}) \times p_{SR}(s, r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{SR}(s, r|Ab)\,. \tag{3.5}$$

The class-conditional probabilities density function can be written as product of conditional and marginal distributions as follows :

$$p_{SR}(s, r|\overline{Ab}) = p_{S|R}(s|R = r, \overline{Ab}) \times p_R(r|\overline{Ab}) \tag{3.6}$$

and

$$p_{SR}(s, r|Ab) = p_{S|R}(s|R = r, Ab) \times p_R(r|Ab) \tag{3.7}$$

Once the two class-conditional probabilities are modeled for each (s,r) value, the problem was to decide if there was enough information to stop reading ("abandon" class), or to continue reading ("no abandon" class). Then, we had to estimate the prior probabilities such as: $P(Ab) + P(\overline{Ab}) = 1$ in order to find the decision rule. Finally, substituting Eqs. 3.6 and 3.7 in 3.5, the decision rule is then:

$$P(\overline{Ab}) \times p_{S|R}(s|R = r, \overline{Ab}) \times p_R(r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{S|R}(s|R = r, Ab) \times p_R(r|Ab) \tag{3.8}$$

3.6.3 Model Learning

Our method was to design a model and to learn its parameters from two-thirds of the data, then test it on the remaining one third of the data. After learning we got a frontier equation. That equation was included in the computational model. That model constantly computes the *Sim* value while it is moving forward in the text, increasing the *Rank* value. As soon as the current *Sim* value observed in the participant's scanpath is greater than the value computed by the model Sim_0 (value in the frontier), the decision is to stop reading the paragraph. Once the model parameters have been set, the evaluation of the model was performed and now it is explained.

3.6.4 Model Performance

In order to test the model, we ran it on the remaining one third of the data. For each fixation in this testing set, the model decides either to leave or not to leave the paragraph. If the model did not leave at the time the participant stopped reading, simulation is pursued with the next *Rank* and with the same value of the *Sim*, and so on until the decision is made. In this particular case, our model is less reactive than participant and because there are no more fixations on the participant scapanth, we used the same value of the *Sim*. However, we thought that if the participant would make more fixations, the *Sim* value after consider these extra fixations will be different but for our modeling work, we are not able to predict which words would be fixated. Then, we made a distinction between these two cases: 1) model stops reading either before or at the same time as participant and 2) model stops reading after the participant.

The average difference between the ranks at which model and participant stopped reading was only computed for the case when model stops reading either before or at the same time of participant. For example Fig. 3.8 shows an experimental scanpath with its corresponding model evaluation. In the paragraph, the participant made only 16 fixations and the model made 17 fixations. Then, in such example, the difference between the rank at which model and participant stopped reading was -1 fixation. Clearly here the model seems to be less reactive than participant because it made one more fixation. So, this case is not considered to compute the average difference between the ranks at which model and participant stopped reading because participant decided to stop reading before the model and there are no more fixations to be treated by the model.

To assess the significance of that value, we built a random model which stops reading after each fixation in the testing set with probability p. If it did not abandon, it considers the next fixation and again decides to leave with probability p, etc. We searched for the best p that minimized the average difference with participant ranks. In a similar way, Fig. 3.8 illustrates the comparison between the random model (13 fixations) and participant's data (16 fixations) and it is of 3 fixations. In this case, the

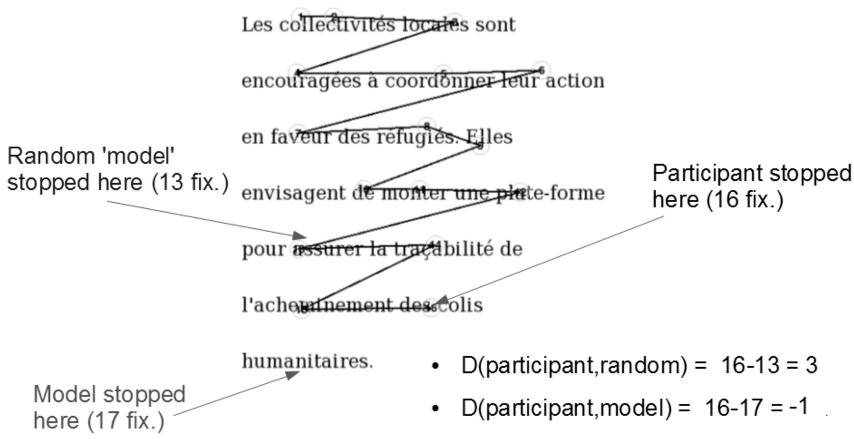


Figure 3.8: Comparison of performance (difference in number of fixations) between participant's data, model and a random model.

random model seems to be more reactive than participant.

After having presented the general methodology that we followed for the development of the two models discussed in the current thesis, we are going to move to the detailed description of our modeling work for each of the experiments on the next two chapters.

Chapter 4

Computational Cognitive Modeling: a simple environment.

In this chapter, the experiment carried out in order to collect data of people searching for information in a simple environment is presented. The experiment corresponds to the Situation 1 (p. 40, Fig. 3.1) that we aim at modeling: deciding as quick as possible if a paragraph presented is related or not to a search goal given beforehand. Our goal was therefore to build a model that would describe the cognitive processes involved in this decision-making situation but also predict when an average user is likely to stop reading the paragraph when a goal is given. As we have also largely discussed in Chapter 2, this particular problem has been studied by Lee & Corlett (2003) where participants were provided with a topic and a text, presented one word every second, and were asked to decide as quickly as possible if the text is about the given topic. However, we aim at studying a normal reading situation instead of presenting one word at a time and because the idea is to model the cognitive processes involved in such an information search task by replicating the human eye movements recorded during a behavioral experiment, we relied on an eye-tracker experiment to be able to identify the words processed.

Along the chapter, we started by giving the description of the experiment carried out. Then model learning and model performance are also discussed. Finally, and for improvement purposes of the model, some of the additional analyses done by following different approaches are also presented.

4.1 Experiment I: one-paragraph experiment

This experiment was intended to emphasize the decision to stop reading a paragraph when participants had to decide as fast as possible whether it was related or not to a semantic goal given at a prior stage.

4.1.1 Participants

A total of 19 French native speakers participated in the experiment (age range 19-43 years, mean age 27 years, sd = 8 years). All participants had normal or corrected-to-normal visual acuity, and had no known neurological disorders. They were naive with respect to the purpose of the study, and all gave written consent.

4.1.2 Material

The material described in Chapter 3 (pag. 47) was used in the experiment: thirty goals with two highly related (HR), two moderately related (MR) and two unrelated (UR) paragraphs.

4.1.3 Apparatus

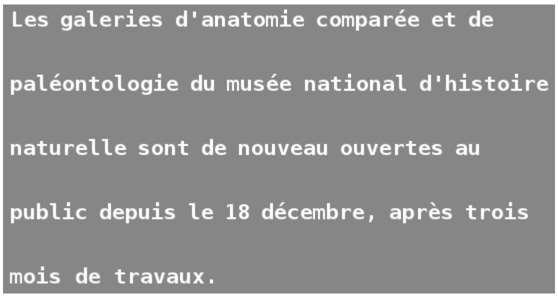
The description of the eye-tracker device is given in Chapter 3 (pag. 48). The EyeLink system was used in the Pupil-Corneal Reflection tracking mode sampling at 1000Hz.

4.1.4 Procedure

The experiment is composed of 180 trials, each one corresponding to a goal (30) with its corresponding 6 paragraphs, in random order. The procedure began with a learning trial using one new goal and six paragraphs. At the beginning of each trial, the goal is displayed. Then a fixation cross is shown and located at the same place where the first character of the first word will appear. The fixation cross is shown, waiting for the participant's gaze to be inside a box of 120 pixels wide and 180 pixels high and centered on the fixation cross. Once the participant's gaze is detected inside the box, the gaze position was controlled to remain inside during more than a random value normally distributed (mean=800 ms, σ=40) and truncated between 700 ms and 900 ms. After this gaze stabilization period, a paragraph is displayed. Participants were asked to decide as quick as possible if the current paragraph is related or not to the goal shown at the beginning. For example, suppose that the paragraph of the Fig. 4.1 is displayed, and we should decide if it is related or not to the goal "art contemporain". What will be the answer? Making a decision in this case is not easy because the relatedness between the paragraph and the goal is not clear. During the experiment, the participants had to click on the mouse only to indicate that the decision had been taken. Immediately, a new screen is displayed in order to collect the participant's decision: to *keep* or to *reject* the paragraph. Participant should *keep* it if according his judgment the paragraph is related to the goal and he should *reject* it in other case.

The participant knew that between each goal, a pause could be managed if necessary and he/she knew the number of remaining goals until the end. A screen indicated that a new goal would be presented as well as the count of the remaining goals. Figure 4.2 shows the different screens presented to the participant during a complete trial.

A 9-point calibration routine was automatically done if the timeout on the initial fixation cross elapsed or if the experimenter decided to run it. The idea behind of the



Les galeries d'anatomie comparée et de
paléontologie du musée national d'histoire
naturelle sont de nouveau ouvertes au
public depuis le 18 décembre, après trois
mois de travaux.

Figure 4.1: Example of paragraph for the goal "art contemporain".

use of a random time at detecting the gaze inside the box was for avoiding participant's anticipation of the reading. A drift procedure is done before each trial to ensure the recording accuracy of eye movement. Running a re-calibration step of the eye tracker was also possible if the experimenter found it necessary. During the experiment, participant's head position was stabilized with a chin rest and their forehead supported on a fixed bar. Participants were seated 68 cm in front of a 24-inch monitor (42° x 21° of visual field) with a screen resolution of 1024 by 768 pixels. The text was displayed at the centre of the screen (21° x 11° of visual field). In average, the text is displayed with 40 characters per line, corresponding about 4 characters in fovea. We wrote our experiments in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

4.2 Pre-processing data

All the procedure described in Chapter 3 (pag. 48) was followed in order to adequate eye-tracker data to a more accessible format for our analysis. The pre-processing stage included the following steps:

- Converting a EDF file to an ASC file by using the EDF2ASC converter.
- Generating a fixation file from an ASC file.
- Adding the words actually processed during each fixation to the fixation file by using our window-mechanism.

Once data were adequate, and before to do a deeply analysis, some basic statistics were computed according to the type of paragraphs: HR (Highly Related), MR (Moderately Related) and UR (Unrelated). The description of these statistics are given now.

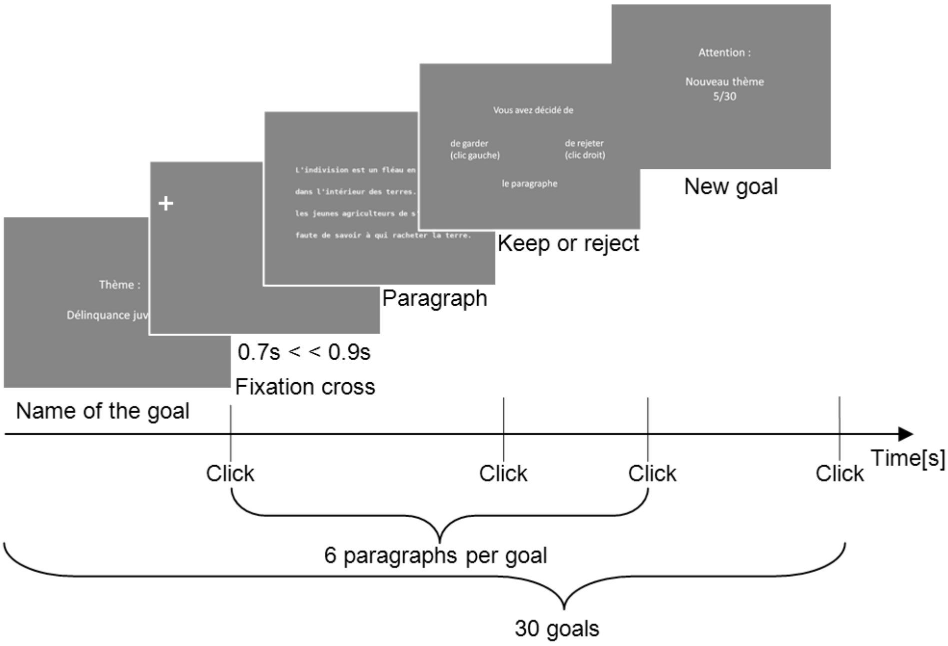


Figure 4.2: Temporal sequence of a trial composed of the different screens of the Experiment I.

4.3 Basic statistics

The basic statistics include the computation of total number of fixations on paragraphs and their mean durations. The ratio of correct responses (*keeping* the paragraphs HR, and *rejecting* the paragraphs UR) was also computed. Table 4.1 shows the mean number of fixations and its mean duration for all the participants according to the nature of paragraphs. The difficulty to decide if a paragraph of type MR is related or not to a given goal is reflected in such Table 4.1. Participants need to do more fixations and the overall fixation duration is also longer on such paragraphs than on paragraphs of type HR and UR. The extra number of fixations is due to the fact that participants can not judge easily about the relatedness (or unrelatedness) of the paragraph. They sometimes decide to *keep* them but in other cases they decide to *reject* them. Their behavior is not easy to predict from just a few words. However, given a goal, decide if a paragraph HR is related to it or if a paragraph UR is not related to it is relatively more easier. Table 4.2 shows the ratio of correct responses. As it was expected, the ratios for both HR (*keeping* the paragraph) and UR (*rejecting* the paragraph) were high: HR: 92.9% vs 7.1%; UR: 5% vs 95%. That means that participants were able to discern about the no-relatedness of paragraphs UR and the relatedness of paragraphs HR with the corresponding goal. LSA and the participant's judgments were consistent. Ratio

Type of Paragraph	Number of fixations (SD)	Total duration in s (SD)
HR	17.3 (8.8)	3.2 (1.7)
MR	22.1 (9.7)	4.0 (1.9)
UR	16.7 (8.7)	3.0 (1.7)

Table 4.1: Mean number of fixations and its mean duration for all the participants according to the types of paragraphs.

Type of Paragraph	Keep (%)	Reject (%)
HR	92.9%	7,1%
MR	47.2%	52.8%
UR	5%	95%

Table 4.2: Ratio of the correct responses for all the participants according to the types of paragraphs.

of correct responses for MR paragraphs are much more balanced (47.2% vs 52.8%), a situation that confirms their role of neutrality and difficulty to discern about their relatedness with the corresponding search goal.

So far, the experiment, the preprocessing stage and the computation of some basic statistics have been explained. Now, we turn into the modeling section.

4.4 Modeling

Making predictions in processing a paragraph. The model should be able to predict the way a paragraph is processed given a goal. For example, given the paragraph and goal of Fig. 4.3, the model should be able to predict the way an average user would process the paragraph. The paragraph might be processed partially and abandoned after about a dozen words have been fixated as illustrated in such example.

Modeling Semantic Judgments. Such a decision making model on paragraphs needs to be based on a model of semantic memory that would be able to mimic human judgments of semantic associations. The robustness of LSA (Landauer and Dutnais, 1997; Landauer et al., 2007) to capture the semantic association between words has been largely demonstrated during the years and at least empirically we can observe roughly that on the computation of the ratio of correct responses in our experiment, LSA and the participant's judgments were consistent. As it was described in Chapter 3 (pag. 47), LSA was used to control globally the relatedness of the paragraphs (material) with the corresponding goals, but here LSA is used at the level of words fixated. Then, we used LSA to dynamically compute the semantic similarities between the goal and each

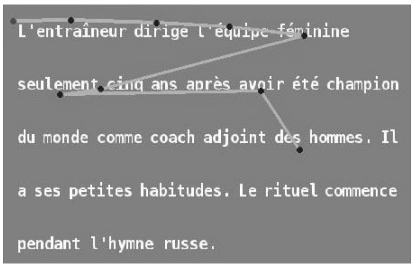


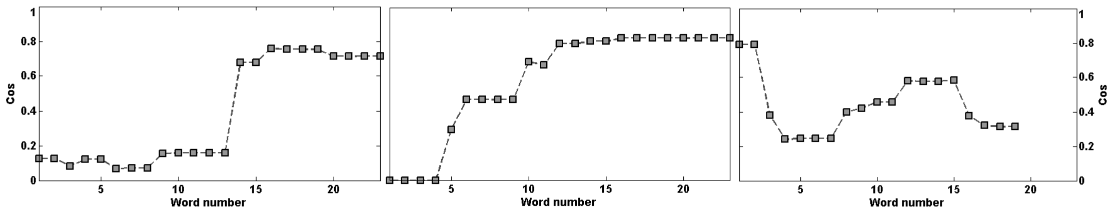
Figure 4.3: Example of scanpath of the Experiment I.

set of words that are supposed to have been fixated.

Selection of the material. In order to facilitate the design of the model, we only analyzed paragraphs which were highly related (HR) to the goal. The reason is that deciding that a paragraph is related to the goal is probably different from deciding that it has nothing to do with the goal. In the first case, the participant is probably expecting highly-related words whereas in the second case it is the absence of such words as well as the presence of specific unrelated words that would trigger the decision. All paragraphs were intentionally built without additional constraints like the frequency of words, the existence of special words in the paragraph or similar but their length and type of overall similarity with the goal. We therefore ended up with a large variety of patterns of goal-similarity evolution across paragraphs. The goal-similarity evolution curves can be easily calculated dynamically, after each word of the paragraph has been processed with the following expression:

$$Cos_i = |sim(w,\, g)|\,. \tag{4.1}$$

in which sim is the LSA cosine between the two LSA vectors of words of w and of words of the goal g. Given a paragraph w of n words and a goal g, we proceed as follows: first, we compute the LSA cosine similarity between the first word w_1 of the paragraph and the goal given us its corresponding cosine value that is Cos_1. Then, the cosine similarity between the concatenation of the first two words w_1 and w_2 of the paragraph and the goal is computed given us a new cosine value, Cos_2. The new computation of the cosine similarity is between the concatenation of the first three words w_1, w_2 and w_3 of the paragraph and the goal and so on. The procedure is followed until we have concatenated all the n-words $(w_1, w_2 ... w_n)$ of the paragraph and we compute its cosine similarity with the goal Cos_n. The cosine similarity value changes constantly while the paragraph is processed given us an evolution curve such as the ones shown in Fig. 4.4. In some paragraphs the similarity between the goal and the words seen so far is constantly increasing throughout the paragraph, whereas in others there are high peaks of similarity because of highly related words, but in-between the similarity between the



(a) Step (b) Ramp (c) Saw

Figure 4.4: Example of the material-based cosine evolution for HR paragraphs after clustering.

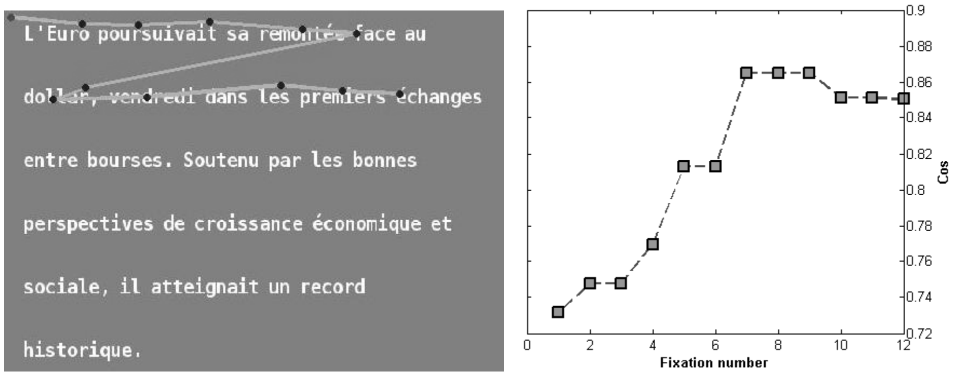
goal and the words seen so far is lower because of general or polysemous words. In order to reduce that variability and simplify the design of the model, we conducted a hierarchical cluster analysis (HCA) of those goal-similarity evolution curves. We found three main kinds of patterns: *Step*, *Ramp* and *Saw*.

The cosine evolution of *Step* paragraphs is characterised by a constant low value during some fixations followed by an abrupt jump to a higher value which is maintained until the end. In the case of *Ramp* paragraphs, a low cosine value is maintained during some fixations but here it is followed by a progressive increasing of the cosine to a higher value which is kept until the end. The cosine evolution in *Saw* paragraphs is still more complicated. Here, the evolution of the cosine value follows a increasing/decreasing pattern along the duration of the signal. We called them *Saw* because of its resemblance with the tool. Figure 4.4 shows an example of a prototype for each cluster.

Of the total number of HR paragraphs (60), some of them were not clustered (23) in any of these classes due the high complexity of the shape of their cosine evolution. For our modeling purposes, only *Ramp* paragraphs (6) were chosen because they have a higher cosine variability than *Step* paragraphs (13) but their variability is more predictable (a nice relationship between its cosine and the word rank as is shown in Fig. 4.4) than the case of *Saw* paragraphs (18). We thought that these particular features may facilitate the modeling work.

4.4.1 Modeling the decision

Two variables involved. As it was discussed before in Chapter 3 (pag. 52), we investigated which variables could play a role in the decision to stop reading a paragraph p, when people were asked to quickly decide if it is related or not to a given goal. We mentioned that the semantic similarity Sim between the words processed in such paragraph and the goal must be involved in such decision. Only the words actually processed at each fixation made by the participant are taken into account. This is not the case of the computation of the material-based evolution curves where all the words



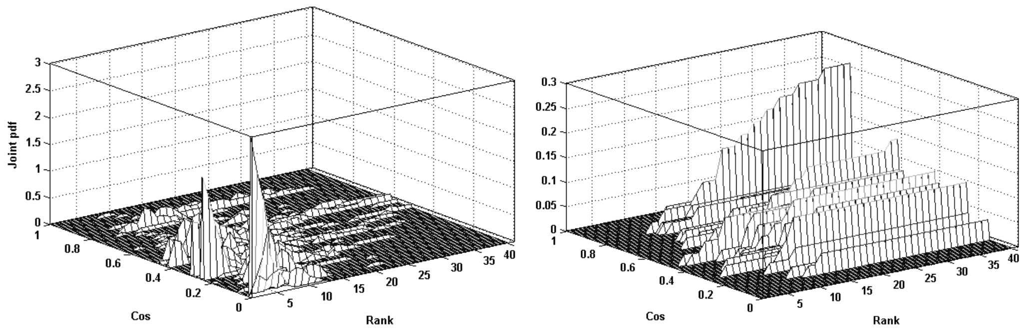
(a) (b)

Figure 4.5: a) Example of scanpath for the goal "faiblesse du dollar". b) Its corresponding participant-based Cos evolution.

contained in a paragraph are considered (in Eq. 4.1). We refer here to such variable Sim as Cos and is defined as follows:

$$Cos = |sim(\text{words actually processed of } p,\, g)| \tag{4.2}$$

in which sim is the LSA cosine between the two LSA vectors of words actually processed of p and of words of the goal g. Cos changes constantly while processing a paragraph. When the words seen are highly related to the goal, Cos has a high value ($\sim$1) and it has a low value ($\sim$0) when the words are unrelated with the goal. It can also be easily calculated dynamically, after each word of the paragraph has been processed. We could call to the resulting curve as the participant-based cosine evolution. Consider for example Fig. 4.5a. The goal is *"faiblesse du dollar"* (*"weakness of the dollar"*). In the first fixation on the paragraph, only the words *"L'Euro"* are supposed to have been processed according to our window-based prediction (Ch. 3, pag. 47). These words are highly related with the goal. Therefore $Cos = |sim($*"L'Euro"*,*"faiblesse du dollar"*$)| = 0.731$. During fixation 2, two words are processed, the word *"L'Euro"* and a new word *"poursuivait"* leading to a new value of $Cos = |sim($*"L'Euro poursuivait"*,*"faiblesse du dollar"*$)| = 0.747$. During fixation 3, only the word *"poursuivait"* is processed, leading to the same value of $Cos = 0.747$. In fixation number 4, $Cos = |sim($*"L'Euro poursuivait sa remontée"*,*"faiblesse du dollar"*$)| = 0.769$. In fixation 5, the Cos value goes up to 0.812 because of the word *"face"* which makes the LSA vector much more similar to the goal vector. This value is maintained for the fixation 6. In fixation 7 and 8 the Cos value goes up to 0.865 because of the word *"dollar"* which makes the LSA vector much more similar to the goal vector. During fixation number 10, the words *"les premiers"* makes the LSA vector less similar to the goal vector and this effect is



(a) (b)

Figure 4.6: a) Empirical "no-abandon" distribution $\hat{p}_{CR}(c, r|\overline{Ab})$ and b) "abandon" distribution $\hat{p}_{CR}(c, r|Ab)$ in the $Cos \times Rank$ space.

showed by the decreasing value of $Cos = 0.850$. This value is maintained until the 12^{th} fixation after which the participant decided to stop reading the paragraph and indicate the decision. Figure 4.5b shows the participant-based evolution of the Cos value along the fixations in the scanpath.

This example illustrates that a high value of Cos may not directly induce the decision, in particular if it appears too early in the scanpath. As it was also discussed on the previous chapter, we assumed therefore that the decision also depends of a second variable: the *Rank* of fixation which give us an idea about the number of words processed so far in the paragraph.

Once we have determined the two variables that could played the role in the decision to stop reading a paragraph, two empirical fixation distributions were computed in order to study how these variables could trigger the decision. Their descriptions are given now.

Abandon and no-abandon distributions. The two distributions of fixations for the no-abandon and the abandon cases were computed in the $Cos \times Rank$ space of participant data in order to learn the frontier between both cases. To compute the distributions, each participant fixation was associated to a point in the $Cos \times Rank$ space. *Rank* is a discrete measure between 1 and the maximum number of fixations in the data (60 in our case). Cos has been computed according to the formula 4.2, taking into account the words already processed in each paragraph as well as the goal and discretized into one of 100 bins, from 0 to 1. The no-abandon distribution was computed by simply counting the number of fixations that did not lead to an abandon for each cell of the $Cos \times Rank$ grid. It concerns all fixations except the last one of each scanpath. The abandon distribution was built from all very last fixations of all scanpaths, including also subsequent ranks. These fixations can be called virtual fixations. For example, if a given participant on a given text made 13 fixations, the first 12 were counted in the no-abandon distribution and the 13^{th} was counted in the abandon distribution. All virtual fixations from 14 to 60, with the same Cos value as the 13^{th} were also counted in the abandon distribution, because if the participant stopped reading at fixation 13, he would have also stopped at fixation 14, 15, etc. The frontier between these two behaviors (continue or stop reading) is a curve in the $Cos \times Rank$ space. Ideally, that curve would be the intersection between the increasing abandon distribution (it is increasing along the $Rank$ axis because the chances of abandoning is higher and higher while the paragraph is processed) and the decreasing no-abandon distribution.

According to the location of a given a observation (c,r) above or under the curve, the reader's behavior can be predicted. For instance, suppose our model makes its 5^{th} fixation on a word and the similarity between all words processed so far and the goal is 0.412. According to the location of the $(5, 0.412)$ point with respect to the frontier curve, our model is able to decide whether to abandon or not.

As it was already mentioned in Chapter 3, a methodology based on a Bayesian classifier was followed and outlined by Equations 3.3, 3.4, 3.5, 3.6, 3.7 and 3.8 (Ch. 3, p. 54). Now by replacing Sim with the corresponding variable Cos presented in the Eq. 4.2 in the Equations 3.5, 3.6, 3.7 and 3.8, we have the Bayesian classifier (Eq. 4.3), the two class-conditional probabilities density functions (Eqs. 4.4 and 4.5) and the decision rule (Eq. 4.6) as follows:

$$P(\overline{Ab}) \times p_{CR}(c, r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{CR}(c, r|Ab)\,. \tag{4.3}$$

$$p_{CR}(c, r|\overline{Ab}) = p_{C|R}(c|R = r, \overline{Ab}) \times p_R(r|\overline{Ab}) \tag{4.4}$$

and

$$p_{CR}(c, r|Ab) = p_{C|R}(c|R = r, Ab) \times p_R(r|Ab) \tag{4.5}$$

$$P(\overline{Ab}) \times p_{C|R}(c|R = r, \overline{Ab}) \times p_R(r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{C|R}(c|R = r, Ab) \times p_R(r|Ab) \tag{4.6}$$

Now the statistical model to estimate the density functions and the prior probabilities are explained in order to use the classifier.

Parametric model for the "no-abandon" distribution. Figure 4.7 (top, left) shows the empirical marginal distribution $\hat{p}_R(r|\overline{Ab})$. As the $Rank$ increases, the probability of not abandoning the paragraph decreases. This evolution was modeled with a sigmoid function:

$$\varphi(r) = \frac{P_{RMax} \times (1 + e^{-\alpha r_0})}{1 + e^{\alpha(r - r_0)}}\,. \tag{4.7}$$

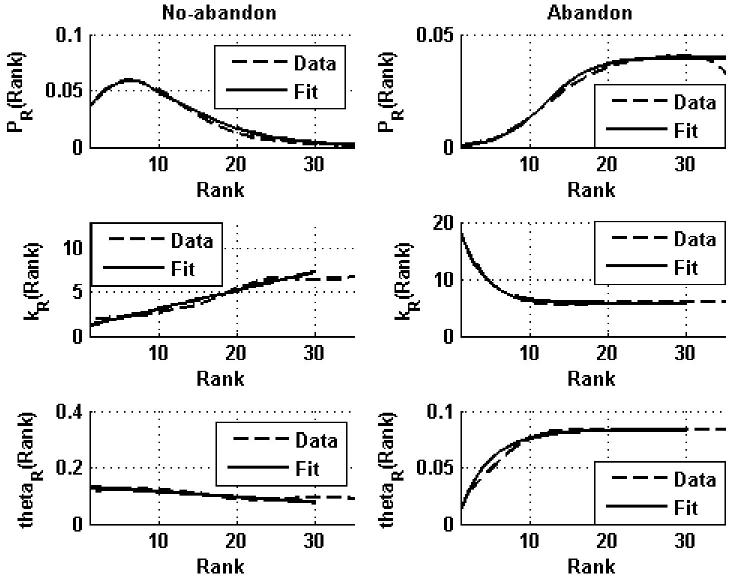
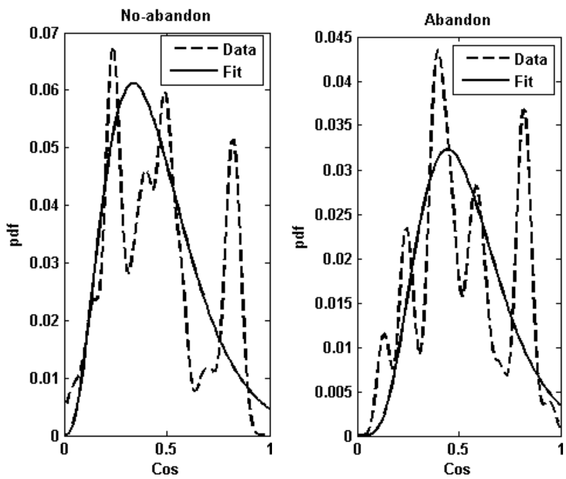


Figure 4.7: Data and fitting of marginal distributions, shape and scale for the "no-abandon" and "abandon" distributions.



(a) (b)

Figure 4.8: Example of empirical conditional pdf and fit with a gamma model for *Rank* value at 10 for the a) "no-abandon" and b) "abandon" class.

There were actually only two parameters r_0 and α to fit because the integral is 1. Concerning the probability density function $p_{C|R}(.)$, the natural model (Fig. 4.6a) is a Gamma one whose parameters depend on the *Rank* value. Figure 4.8a illustrates for a particular *Rank*=10 the reason why the Gamma is a good function to model these data. The shape $k(r)$ increases and the scale $\theta(r)$ decreases linearly (Fig. 4.7, left column). Both are modeled by simple linear equations: $k(r) = Mr + N$ and $\theta(r) = Or + P$. The linear regressions are only performed up to the $Rank = 30$ since that $\hat{p}_R(r > 30|\overline{Ab})$ is close to zero and there is no more enough data. Then we have:

$$p_{C|R}(c|R = r, \overline{Ab}) = \frac{A(r)c^{k(r)-1}}{(k(r)-1)!\theta(r)^{k(r)}} e^{-\frac{c}{\theta(r)}}, p_R(r, \overline{Ab}) = \varphi(r)\,. \tag{4.8}$$

As the *Cos* value is between 0 and 1, $A(r)$ is a normalization function to ensure that $p_{C|R}(c|R = r, \overline{Ab})$ is a probability density function whose integral equals 1. We get $A(r) = F_{k,\theta}(1) - F_{k,\theta}(0)$, with $F_{k,\theta}(.)$ being the repartition function of a Gamma distribution with a shape k and a scale θ. We then obtained six independent parameters to model the complete "no-abandon" joint distribution: offset r_0 and slope α for $\varphi(r)$, the coefficients M,N for $k(r)$ and O,P for $\theta(r)$.

Parametric model for the "abandon" distribution. Following a similar approach the marginal pdf $\hat{p}_R(r|Ab)$ was modeled with another sigmoid function $\varphi'(r)$ (Fig. 4.7, top right). But here, it is an increasing function. At rank 0, there is no abandon and at the maximal *Rank* value, all scanpaths have shown an abandon. The conditional distribution $\hat{p}_{C|R}(c|R = r, Ab)$ is a Gamma distribution with a shape $k'(r)$ and a scale $\theta'(r)$. As in the previous case, Fig. 4.8b illustrates how these data are modeled by a Gamma function. The shape $k'(r)$ exponentially decreases while the scale $\theta'(r)$ exponentially increases (Fig. 4.7, right column). Both are modeled by exponential equations: $k'(r) = (s_2 - s_1)e^{-r/p_0} + s_2$ and $\theta'(r) = (t_2 - t_1)e^{-r/q_0} + t_2$. Then, the different set of functions $\{\varphi'(r),k'(r),\theta'(r)\}$ gives us eight parameters: r'_0 and α' for the $\varphi'(r)$; p_0, s_1 and s_2 for $k'(r)$; q_0, t_1 and t_2 for $\theta'(r)$. The resulting equation for the pdf is:

$$p_{C|R}(c|R = r, Ab) = \frac{A'(r)c^{k'(r)-1}}{(k'(r)-1)!\theta'(r)^{k'(r)}} e^{-\frac{c}{\theta'(r)}}, p_R(r, Ab) = \varphi'(r)\,. \tag{4.9}$$

Table 4.3 summarizes the total number of parameters and their values for both modeled distributions: the no-abandon and the abandon.

4.4.2 Model Learning

Once the model was designed, its parameters were learned from two-thirds of the data. The two prior probabilities $P(Ab)$ and $P(\overline{Ab})$ were also learned by considering how

Distribution	Parameters
No-abandon	r_0=11.97, α=0.17, M=0.21, N=1.0, O=-0.0018, P=0.13
Abandon	r'_0=12.0, α'=0.33, $p_0 = 2.94, s_1$=5.76, s_2=23.70, q_0=3.85, t_1=0.083 t_2=-0.001

Table 4.3: Model parameters for both the no-abandon and the abandon distributions.

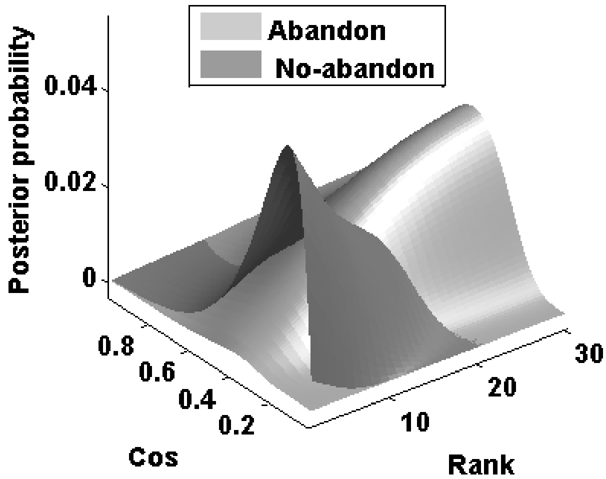


Figure 4.9: The posterior prob. $P(\overline{Ab}|c, r)$ and $P(Ab|c, r)$ in the $Cos \times Rank$ space.

much data were of both distributions. The number of occurrences of the abandon distribution was divided by the total number of occurrences of both the abandon and the no-abandon distributions, given us $P(Ab) = 0.64$ and $P(\overline{Ab}) = 0.36$. Then, the total number of learned parameters was 15 (6+8+1).

Figure 4.9 shows the two posterior probabilities $P(\overline{Ab}|c, r)$ and $P(Ab|c, r)$ after learning in order to represent the decision frontier between the two classes. As Fig. 4.9 shows, the intersection is oblique which is what was expected, from a cognitive point of view. *Rank* and *Cos* are dependent on each other: at the beginning of processing the paragraph (low values of the *Rank*), there should be a high relatedness (a high *Cos*) between the paragraph and the goal to make the decision. However, after more fixations have been made, that relatedness could be lower to decide to abandon the paragraph. For instance, at rank 10, a *Cos* of 0.7 is necessary to stop reading, whereas at rank 15, a value of 0.3 is enough. The Bayesian frontier is shown in Fig. 4.10. The frontier does not seem to be represented easily by a basic equation. For that reason, two considerations were made to decide which function should be used to fit the frontier. Firstly, and as we have already mentioned in Chapter 1 (p. 3), we tried to have the minimum number of parameters for the model and to use a simple mathematical equation as possible. The idea was not to fit the frontier perfectly in order to avoid the

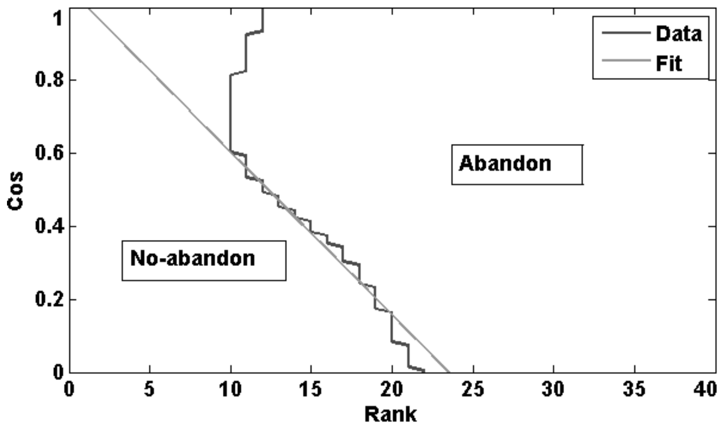


Figure 4.10: The Bayesian frontier between both the abandon and the no-abandon cases. Experimental data and its linear approximation are shown.

noisy data. By this way, our model should be robust enough to make predictions on new data. Secondly, we observed that most of the data are concentrated for *Cos* values below 0.7. This situation can be noticed with the superposition of the densities of both posterior probabilities for the no-abandon and abandon cases as is shown in Fig. 4.11a and 4.11b respectively. Due to that for *Cos* values below 0.7 the frontier seems to be linear, it is approximated by a simple linear equation in the $Cos \times Rank$ space:

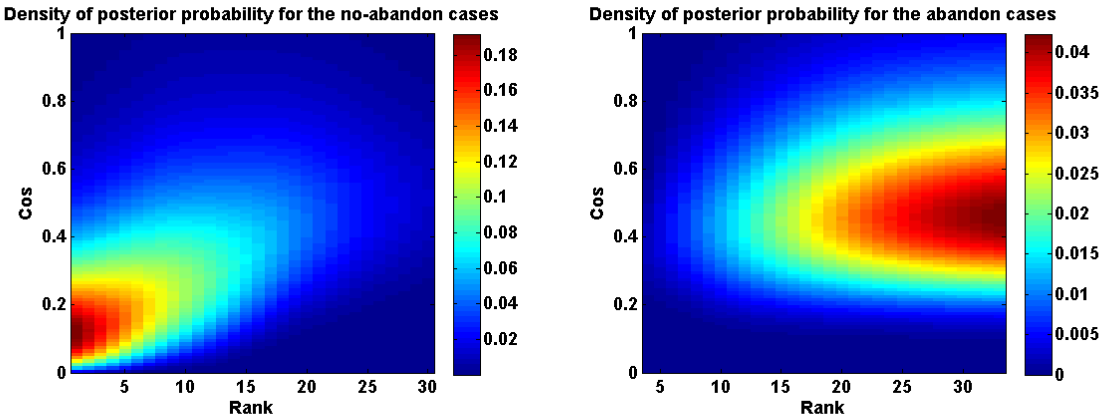
$$Cos_0 = -0.047 \times Rank + 1.05\,.$$

This equation was included in the computational model. That model computes the Cos_0 value while it is moving forward in the text, increasing the *Rank* value. When the current *Cos* value is greater than Cos_0, the decision is to stop reading the paragraph. For example, Fig. 4.12 shows the cosine evolution of an experimental scanpath (dotted line) and the model frontier (straight line). The scanpath is composed of 21 fixations but the model decides to abandon the paragraph at fixation number 17 when the *Cos* value is greater than Cos_0 value computed by the model. However, we can observe that participant made 4 more fixations than model before to stop reading the paragraph.

Once the model parameters have been set, the evaluation of the model is discussed now.

4.4.3 Model Performance

The evaluation of the model performance was done as it was explained on the previous chapter. The model is ran it on the remaining one third of the data. For each fixation in this testing set, the model decides either to leave or not to leave the paragraph. We made a distinction between the two cases: 1) model stops reading either before or at



(a) No-Abandon (b) Abandon

Figure 4.11: Densities of posterior probabilities for the no-abandon and the abandon distributions.

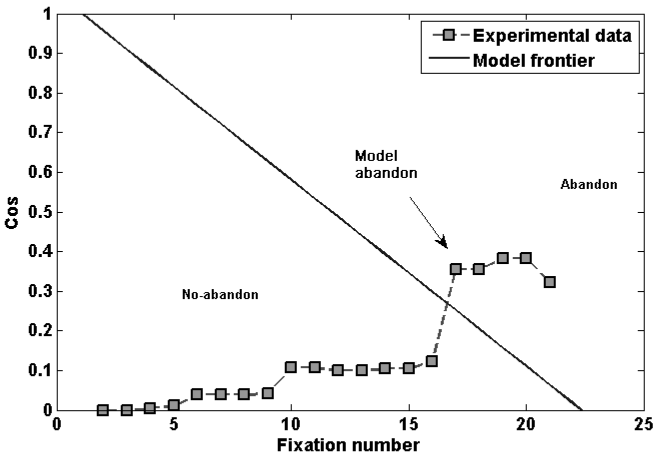


Figure 4.12: Cosine evolution of an experimental scanpath and the model frontier.

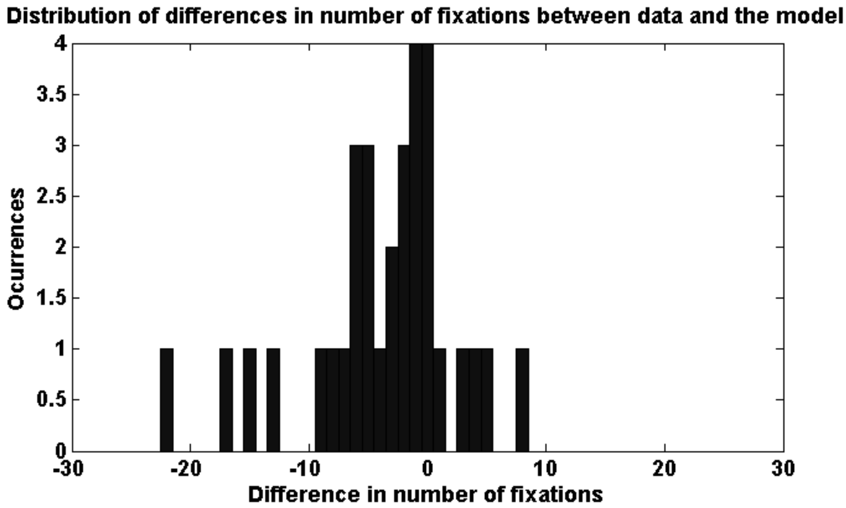


Figure 4.13: Distribution of the differences between the number of fixations (ranks) made by the participant and the model.

the same time of participant and 2) model stops reading after the participant. The case 1 was observed in 81.25% of the trials and in 18.75% of the trials the model stopped reading after the participant.

The average difference between the ranks at which model and participant stopped reading was computed only for the case number 1 where the model is more reactive than participant. We got a value of 4.7 (SE=0.72). It means that, in the average, the model stopped reading 4.7 fixations earlier than participants. The distribution of the differences between the number of fixations made by the participant and the model is showed in Figure 4.13. The difference between the rank at which model and participant stopped reading is indicated on the horizontal axis and the number of occurrences on the vertical axis. A negative value for the difference of ranks (to the left) means that our model is more reactive than participant and viceversa (to the rigth). Zero means that the model is resulting as reactive as participant. A total of 32 cases (scanpaths) were considered in the testing set.

For comparison purposes with the random model built (Ch. 3, pag. 55), the best p value found such as the random model had the smallest average difference with participant ranks was for $p = 0.1$. That smallest average difference was of 11.09 (SE=1.2). So, the results presented in the current section indicates that our model is nearly three times (in average) better than the best random model (4.7 fix. vs 11.09 fix.).

4.5 Conclusion

So far, we have presented a computational cognitive model that is able to replicate the average people's behavior, when they are faced to a task of information search and more specifically when they are asked to decide as soon as possible if a paragraph presented is related or not to a given search goal. The model predicts the sequence of words that are likely to be fixated before the paragraph is abandoned. Two variables seem to play a role: the *Rank* of the fixation and the semantic similarity *Cos* between the words processed so far and the search goal. We proposed a simple linear threshold to account for that decision. The results are promising but still too far of a ideal model, which is not the aim in the current thesis, but we believed in the possibility to improve the results by following different approaches and in the next sections, the description of such approaches is given.

4.6 Improvement

Rank and *Cos* are the variables that play a role in the computational cognitive model presented so far, and here we tried to investigate which other variables or factors would trigger the decision to stop reading a paragraph. There is no doubt that participants make a different number of fixations on paragraphs before to stop reading it and our hypothesis is that participant's idiosyncracy also plays a role. For example some participants are more conservatives than others. The more conservatives might need to do more fixations than the less conservatives even if they had been realized about the relatedness or unrelatedness of a paragraph with the goal. These extra fixations made for each participant can be seen as a necessity of a confirmation stage which is variable across participants. Then the idea was to characterise each participant with such a value (eg. number of extra fixations) and we called it participant's delay or participant's signature.

Then the problem addressed was how to compute the participant's delay. We have followed two different approaches. In our first approach, we suspected that fixations on *special* words would play a role in triggering the decision to stop reading a paragraph. Saying differently, our hypothesis here was that when a *special* word is fixated, the number of fixations needed to stop reading the paragraph is modulated by the nature of such *special* word. Several *special* words were considered and their description will be given later. In this approach, the participant's delays were learned as independent values from all the other participants. In the second approach the participant's delays were learned by comparing each participant's distribution with the distribution of a participant as reference and by computing the Kullback-Leibler (KL) divergence. Here each participant's delay depends of the participant reference. This approach is also explained later.

The ultimate goal is to use the participant's delays to improve the model performance by using these values for the modeling work as follows: given a participant with a delay d and a valid scanpath of n fixations, only the first n-d fixations will be considered for that particular scanpath because the remaining d fixations are considered that were made during the confirmation stage of the participant and the decision was taken not after the fixation $(n-d)$. When the delay value d is either equal or greater to the number of fixations n of a participant's scanpath, the scanpath is totally excluded for the modeling work. When $d = 0$, all the n fixations of the scanpath would be considered. The procedure is done for all the scapanths of the corresponding participant.

Now, we will start by explaining how the participant's delays were learned by using our first approach and then their application to our modeling work is also discussed.

4.6.1 Participant's delays learned by using the *step* word, *target* word and the *ramp* word as references

As in the previous sections, only the paragraphs with a strong semantic association with the search goal (HR) were considered. Additionally, paragraphs for which the evolution of their semantic similarity with the goal follows a *step* or a *ramp* shape were only studied. Therefore, the analysis only concerns paragraphs HR which follow *step* or *ramp* shapes called "HR-*step*" and "HR-*ramp*" respectively. The first two words considered as candidates to be *special* were the *step* and the *target* words. Additionally, a *ramp* word was also considered. Their descriptions are now given.

Given a set of fixations on a paragraph "HR-*step*", the *step* word represents the word which makes that the semantic similarity between the words seen so far and the goal increases from a very low value to a higher value and this value is maintained till the end of the fixations for this particular scanpath. For example, on Figure 4.4a the 15^{th} fixation is made on the *step* word because it is there where the signal shows an abrupt jump to a higher value which is maintained until the end of the scanpath.

A *target* word is in any verbal form of one of the words of the search goal (theme). For example, if the search goal is "croissance de l'économie", target words could be: "croissance","économie", "économique", "economistes", "croissante", etc. Of course, the target words are always highly related with the search goal and in general they can be found only in paragraphs of type HR.

In the case of paragraphs "HR-ramp", we also tried to use a *special* word as reference, but here things are different because there are two parameters defining the *ramp*: the beginning word (initial index) and the ending word (final index) of the *ramp* respectively. For example, on Figure 4.4b the 5^{th} and 13^{th} fixations are made on the beginning word and ending word of the *ramp*. We have considered the ending word of the *ramp* as the *special* word and we will refer it only as *ramp* word.

Then, we have predicted the delay value for each participant. This value is their

Participant	Min	Max	Med	Mean	Mode
s1	2	21	7.0	8.9	5.0
s2	2	19	11.0	9.4	11.0
s3	1	22	9.0	8.6	11.0
s4	7	39	29.0	25.3	7.0
s5	1	30	15.0	15.6	1.0
s6	0	28	11.0	11.2	11.0
s7	1	16	7.0	7.2	7.0
s8	2	28	8.0	11.0	8.0
s9	9	39	24.0	21.6	27.0
s10	4	22	10.0	10.2	4.0
s11	0	4	2.0	1.7	1.0
s13	2	24	10.0	12.3	19.0
s14	1	33	7.0	10.1	5.0
s17	2	39	16.0	17.7	5.0
s18	0	21	8.0	8.8	8.0
s19	0	17	8.0	8.6	11.0
s20	2	21	5.0	8.6	2.0
s21	0	21	10.0	10.1	8.0

Table 4.4: Participant's delays learned by using the *step* word as reference.

idiosyncratic tendency to wait after a decision has been made in their mind. To do that, we relied on the *special* words that are likely to trigger the decision: *step*, *target* and *ramp*. We computed the number of additional fixations made after one of these special words were fixated.

For the case of the *step* word, we ended up with the results showed in Table 4.4. The table shows for each participant the minimum (Min), maximum (Max), median (Med), mean (Mean) and mode (Mode) of the number of fixations made after the fixation on the *step* word up to the end of reading. For example, participant "s1" abandoned the paragraph in average 8.9 fixations after that the *step* word was fixated, but there was at least one case where he abandoned the paragraph after 2 fixations (Min) and after 21 fixations (Max). More often, the participant abandoned the paragraph after 5 fixations after the fixation on the *step* word. The median indicates that he abandoned the paragraph 7 fixations after the *step* word. In a similar way for the case of the *target* word as reference, Table 4.5 shows the participant's delays learned. Here, participant "s7" abandoned the paragraph in average 4.6 fixations after that the *target* word was fixated, but there was at least one case where he abandoned the paragraph after 2 fixations (Min) and after 7 fixations (Max). Mode and median indicate that the participant abandoned

Participant	**Min**	**Max**	**Med**	**Mean**	**Mode**
s1	1	13	7.0	6.6	7
s2	1	21	9.0	9.8	2
s3	2	17	5.7	7.4	3
s4	1	24	5	9.2	1
s5	3	22	4.0	9.0	4
s6	1	15	6.0	7.0	1
s7	2	7	5.0	4.6	5
s8	2	17	5.0	7.2	2
s9	4	25	12	12.6	5
s10	3	27	6.0	9.0	5
s11	0	13	5	4.0	1
s13	2	13	5.0	6.0	5
s14	5	21	12.5	12.7	5
s17	1	22	8.0	9.5	2
s18	1	9	4.0	4.3	1
s19	1	14	2.5	4.8	1
s20	1	24	4.0	7.1	1
s21	1	11	3.0	4.27	1

Table 4.5: Participant's delays learned by using the *target* word as reference.

Participant	**Min**	**Max**	**Med**	**Mean**	**Mode**
s1	0	9	2.0	3.6	2.0
s2	0	9	3.0	3.8	3.0
s3	0	5	2.5	2.6	2.0
s4	4	30	13.5	16.0	4.0
s5	1	7	4.0	4.0	2.0
s6	0	8	3.0	3.5	0.0
s7	0	7	3.0	2.8	0.0
s8	1	7	4.5	4.5	4.0
s9	0	8	7.0	5.6	7.0
s10	0	15	1.5	4.0	1.0
s11	0	6	1.0	2.0	1.0
s13	0	10	4.5	4.3	5.0
s14	0	3	1.0	1.2	0.0
s17	0	6	2.0	2.8	2.0
s18	1	7	4.5	4.3	3.0
s19	1	17	2.5	4.8	2.0
s20	0	8	4.0	4.0	4.0
s21	0	6	1.5	2.1	1.0

Table 4.6: Participant's delays learned by using the *ramp* word as reference.

the paragraph 5 fixations after the fixation on the *target* word.

Finally, Table 4.6 shows the participant's delays learned by using the *ramp* word as reference. Here, participant "s21" abandoned the paragraph in average 2.1 fixations after that the *ramp* word was fixated, but there was at least one case where he abandoned the paragraph after any fixation (Min) and after 6 fixations (Max). Mode and median indicate that the participant abandoned the paragraph 1 and 1.5 fixations after the fixation on the *ramp* word.

Fig. 4.14 plots the participant's delays curves learned for all the participants by using the *step*, *ramp* and *target* words as references. The mode value was chosen as participant's delay. In average, the participant's delays learned by using the *step* word are higher (mean=8.3, SE=1.5) than those learned by using the *ramp* (mean=2.3, SE=0.4) and the *target* word (mean=2.8, SE=0.4) and are significantly different. In the case of the values learned by using the *ramp* (mean=2.3, SE=0.4) and the *target* (mean=2.8, SE=0.4) words, there is not a significant difference (t-test, p>0.4) and due their low variability, we suspected that they should be good candidates to be used as participant's delay in our modeling work.

Once the participant's delays were learned and by following the same approach

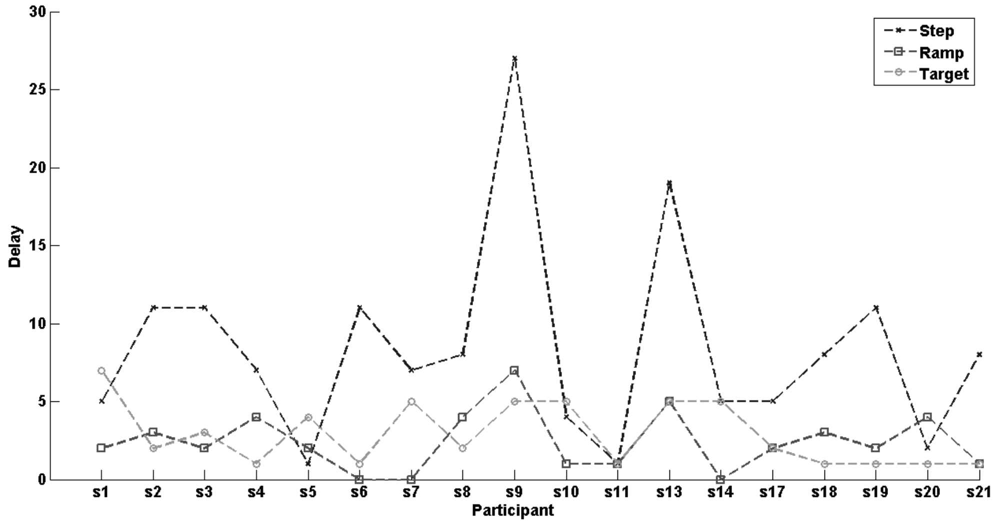


Figure 4.14: Participant's delays learned for all the participants by using the *step*, *ramp* and *target* words as references.

described in Section 4.4.1 (p. 63) to build the model, the two empirical fixation distributions were also computed in the $Cos \times Rank$ space: the no-abandon and the abandon distributions. The idea was to find the frontier between both, but before trying to fit the data, we did a previsualisation of the frontier with a very simple smoothed version of both experimental distributions. In the case of the *step* word the frontier does not seem to be a simple function between *Rank* and *Cos* as Fig. 4.15 shown. In the case of the *target* word the frontier does not seem to be a function with a dependency between *Rank* and *Cos* as is shown in Fig. 4.16. Having a simple function with a dependency between *Rank* and *Cos* would be nicely from a cognitive point of view. To sum up, by using the participant's delays learned either from the *step* or *target* words, the model did not *run*.

However, for the case when participant's delays were learned from *ramp* words things were different. Once these delays were applied, the frontier between both experimental distributions (smoothed versions) showed a dependency between *Rank* and *Cos* as we can see in Fig. 4.17. Then, we proceeded with our modeling work as before. We computed both empirical fixation distributions: the no-abandon and the abandon distributions in the $Cos \times Rank$ space. Both empirical distributions are shown in Figure 4.18. Following a similar approach as in the development of our first version of the model, the Bayesian classifier and a parametric approach to fit both distributions were also used. Equations were the same for the Bayesian classifier (Eq. 4.3, pag. 66), the two class-conditional probabilities density functions (Eqs. 4.4 and 4.5, p. 66) and the

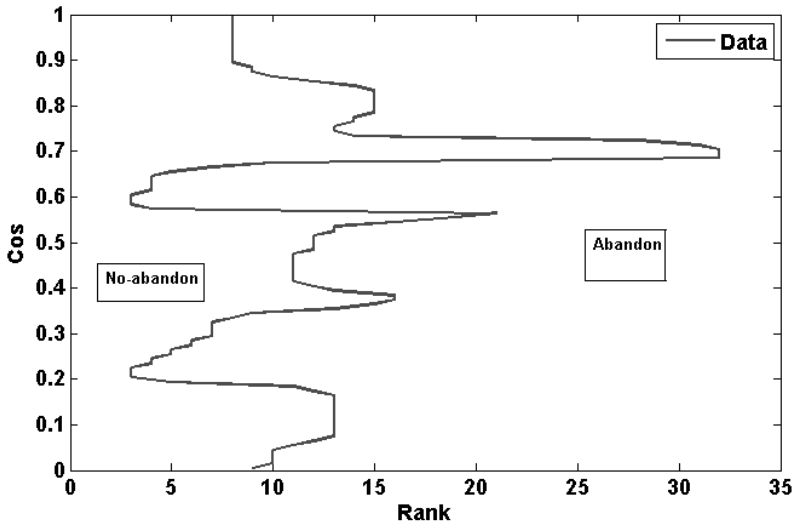


Figure 4.15: The frontier between both the abandon and the no-abandon distributions. The frontier is not a simple function between *Rank* and *Cos*. Delays were learned by using the *step* word as reference.

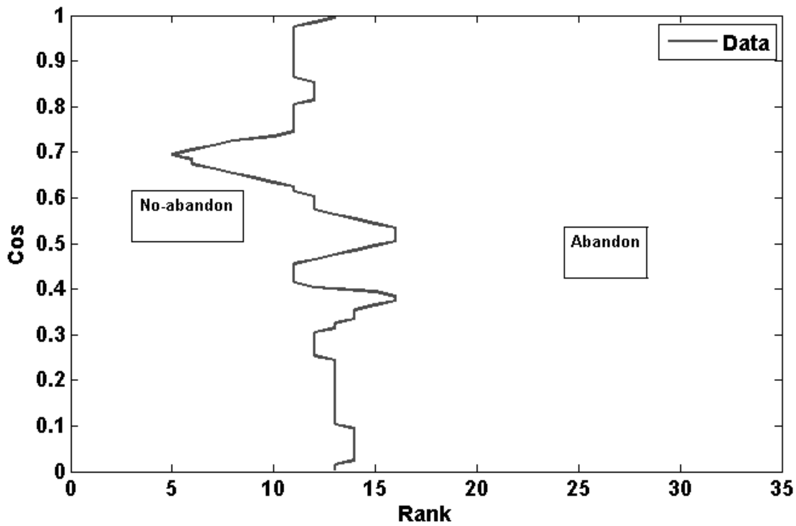


Figure 4.16: The frontier between both the abandon and the no-abandon distributions. The frontier does not seem to be a function with a dependency between *Rank* and *Cos*. Delays were learned by using the the *target* word as reference.

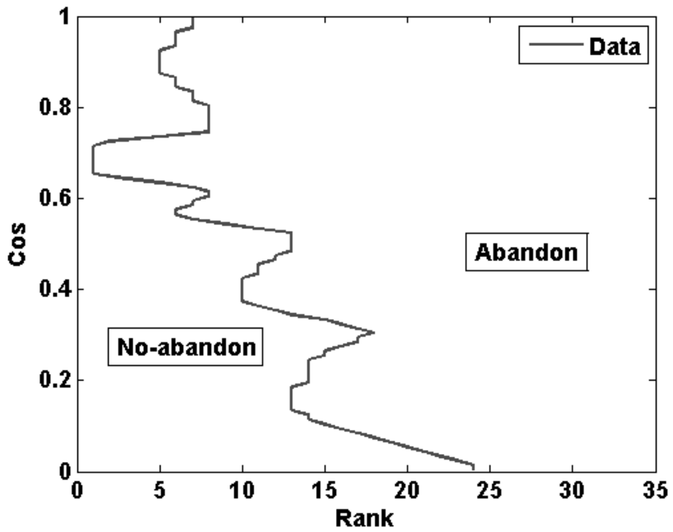


Figure 4.17: The frontier between both the abandon and the no-abandon distributions. The frontier seems to be a function with a dependency between *Rank* and *Cos*. Delays were learned by using the *ramp* word as reference.

decision rule (Eq. 4.6, p. 66). In the case of the no-abandon distribution, Fig. 4.19 (top, left) shows the empirical marginal distribution $\hat{p}_R(r|\overline{Ab})$. As the *Rank* increases, the probability of not abandoning the paragraph decreases. This evolution was also modeled with a sigmoid function (Eq. 4.7, p. 66). Concerning the probability density function $p_{C|R}(.)$, the natural model (Fig. 4.18a) is also a Gamma one whose parameters depend on the *Rank* value. Then we have:

$$p_{C|R}(c|R=r,\overline{Ab}) = \frac{A(r)c^{k(r)-1}}{(k(r)-1)!\theta(r)^{k(r)}}e^{-\frac{c}{\theta(r)}}, p_R(r,\overline{Ab}) = \varphi(r)\,.$$

The shape $k(r)$ increases and the scale $\theta(r)$ decreases squarely (Fig. 4.19, left column). Both are modeled by squared equations: $k(r) = Tr^2+Ur+V$ and $\theta(r) = Wr^2+Xr+Z$. We then obtained eight independent parameters to model the complete "no-abandon" joint distribution: offset r_0 and slope α for the sigmoid $\varphi(r)$, the coefficients T,U,V for $k(r)$ and W,X,Z for $\theta(r)$.

In the case of the abandon distribution, the marginal pdf $\hat{p}_R(r|Ab)$ was modeled with another sigmoid function $\varphi'(r)$ (Fig. 4.19, top right). But here, it is an increasing function. The conditional distribution $\hat{p}_{C|R}(c|R=r,Ab)$ is a Gamma distribution with a shape $k'(r)$ and a scale $\theta'(r)$.

The shape $k'(r)$ decreases while the scale $\theta'(r)$ increases (Fig. 4.19, right column). Both are modeled by two 3-order polynomials: $k'(r) = Gr^3+Hr^2+Ir+J$ and $\theta'(r) =$

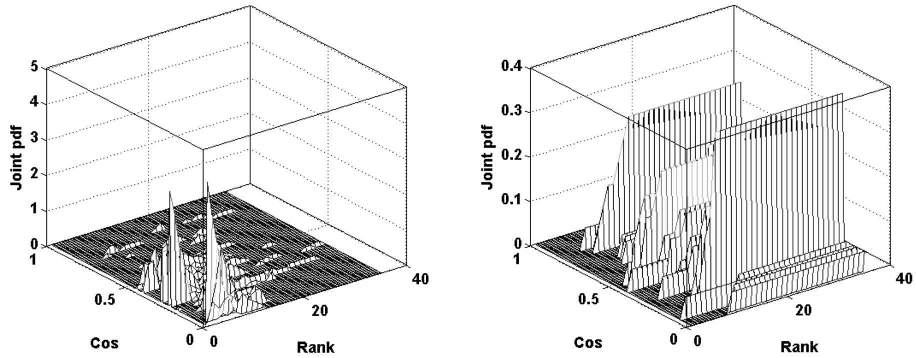
Distribution	Parameters
No-abandon	r_0=11.0, α=0.196, T= 0.0018, U=0.112, V=1.61 W= $3.9140e-005$, X=-0.0022, Z= 0.123
Abandon	r_0'=0.868, α'=0.128, G = -1.363e-004, H=0.0091, I= -0.172 J=5.089, M=3.421e-006, N=-2.324e-004, O= 0.005, P=0.076

Table 4.7: Model parameters for both the no-abandon and the abandon distributions. Participant's delays learned by using the *ramp* word were applied.

Mr^3+Nr^2+Or+P. Equations of the pdf are the same as the previous case, but with a different set of functions $\{\varphi'(r)$,$k'(r)$,$\theta'(r)\}$ which gives us ten parameters (offset r_0 and slope α for the sigmoid $\varphi'(r)$; G,H,I and J for $k'(r)$; M,N,O and P for $\theta'(r)$). The prior probabilities $P(Ab) = 0.67$ and $P(\overline{Ab}) = 0.33$ were learned from data as before. Then, the total number of learned parameters was then 19 (8+10+1). Table 4.7 summarizes the total number of parameters and their values for both modeled distributions: the no-abandon and the abandon.

The intersection between the two modeled distributions $P(\overline{Ab}|c,r)$ and $P(Ab|c,r)$ after learning is oblique as it is shown in Fig. 4.20. The frontier seems to be a linear function as is shown in Fig. 4.21 for *Cos* values below of 0.8 where most of the data are concentrated.

The frontier is rather linear as is shown in Fig. 4.21 and can be approximated by the following equation in the $Cos \times Rank$ space that it is included in the computational model: $Cos_0 = -0.0449 \times Rank + 0.7881$. The distribution of the differences between the number of fixations made by the participant and the model is shown in Figure 4.22. We got as difference of the ranks at which model and participant stopped reading of about 3.0 (SE=0.5) fixations. As before, this computation was made only for the case 1 where the model is more reactive than participant. The result is apparently better than the model's performance presented without any consideration of participant's delay (3.0 fix vs. 4.7 fix). However, we would like to point out that this result should be seen with caution and not as an improvement of our model as we will discuss on the next chapter. The problem is that the *ramp* word is not well-defined in a single point as in the case of *target* or *step* words. A *ramp* is defined by two words: the starting and final words. Any word in between is part of the *ramp* and it could be used as *ramp* word. We have used the word index where the *ramp* ends (final word) but there was no any particular reason to do that.



(a) (b)

Figure 4.18: a) Empirical "no-abandon" distribution $\hat{p}_{CR}(c, r|\overline{Ab})$ and b) "abandon" distribution $\hat{p}_{CR}(c, r|Ab)$ in the *Cos* $\times$ *Rank* space. Participant's delays learned by using the *ramp* word were applied.

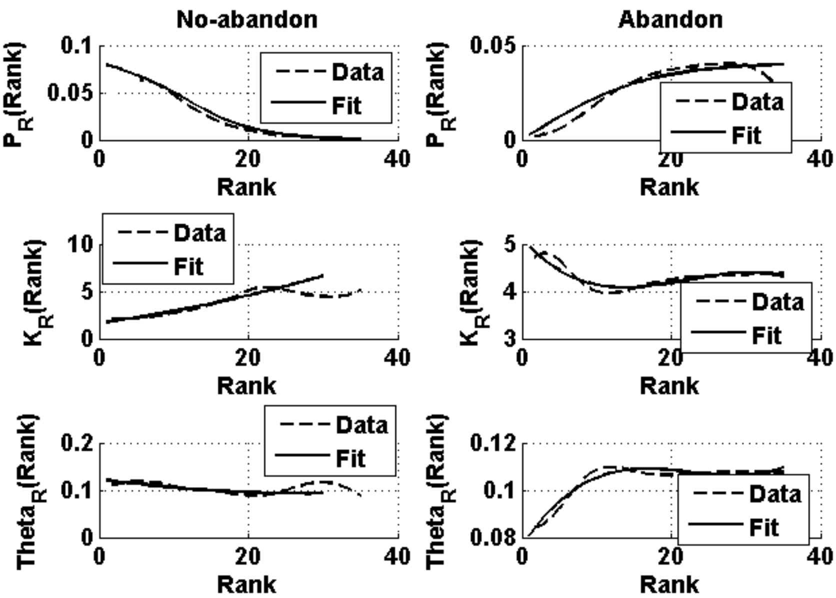


Figure 4.19: Data and fitting of marginal distributions, shape and scale for the "no-abandon" and "abandon" distributions. Participant's delays learned by using the *ramp* word were applied.

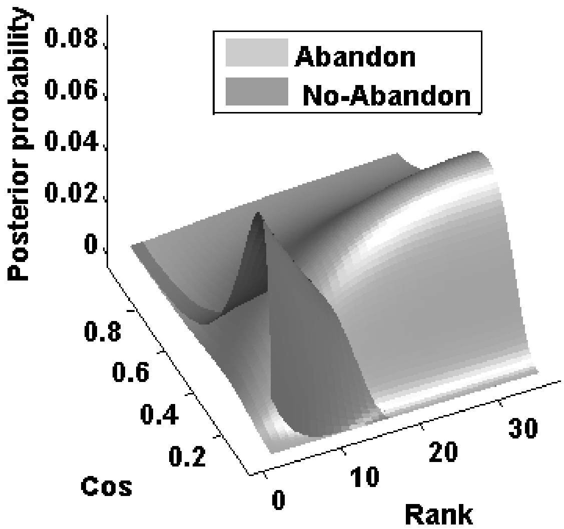


Figure 4.20: The posterior prob. $P(\overline{Ab}|c, r)$ and $P(Ab|c, r)$ in the $Cos \times Rank$ space. Participant's delays learned by using the *ramp* word were applied.

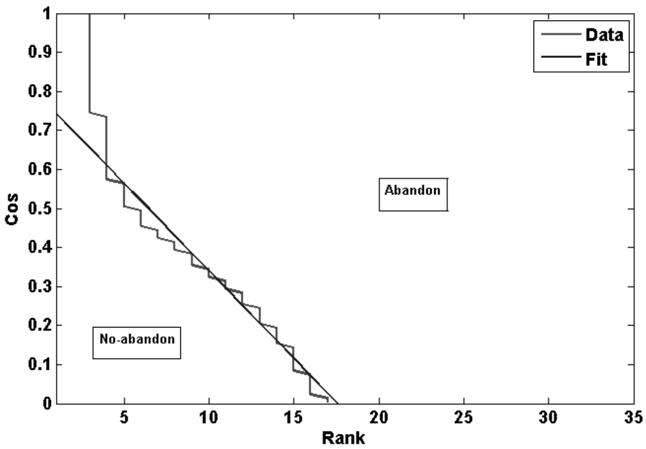


Figure 4.21: The Bayesian frontier between both the abandon and the no-abandon cases. Delays were learned by using the *ramp* word as reference. Experimental data and its linear approximation are shown.

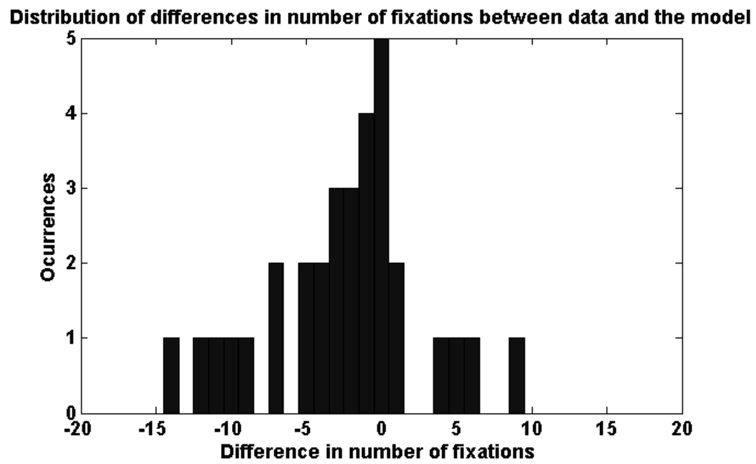


Figure 4.22: Distribution of the differences between the number of fixations (ranks) made by the participant and the model. Delays learned from the *ramp* word were applied.

4.6.2 Learning participant's delays by using the Kullback-Leibler (KL) divergence

On the previous approaches, the participant's delays were learned as independent values from all the other participants (absolute values), but here the idea is different and it can be explained as follows: given the empirical abandon distributions of the participants $\hat{p}_{CR}(c, r|Ab)$, we can consider only the marginal distributions $\hat{p}_R(r|Ab)$ and select one of them as reference $\hat{p}^*_R(r|Ab)$, the marginal distribution of a participant (eg. the distribution of the least/most reactive. Later we will explain the implication of such selection). Then, the marginal distributions of the remaining participants are compared with the reference distribution by using the Kullback-Leibler (KL) divergence. The Kullback-Leibler (KL) or Jensen-Shannon divergence (sometimes called information gain or relative entropy) is a non-symmetric measure of the difference between two probability distributions P_1 and P_2. More specifically, the KL divergence of P_2 from P_1, denoted $KL(P_1, P_2)$, is a measure of the information lost when P_2 is used to approximate P_1. Typically, P_1 represents the true distribution of data, observations (in our case the reference distribution). The measure P_2 represents a description or approximation of P_1 (in our case the remaining participant's distributions). The KL divergence is given by:

$$KL(P_1(x), P_2(x)) = \sum_{i=1}^{M} P_1(x_i).log\frac{P_1(x_i)}{P_2(x_i)}.$$

in which P_1 and P_2 are length-M vectors of probabilities representing the distributions 1 and 2 respectively. Thus, the probability of value $x(i)$ is $P_1(i)$ for distribution 1 and

$P_2(i)$ for distribution 2.

Then, in our case we proceeded as follows: given a participant's distribution (marginal) the idea was to find a shift value (along the *Rank* axis) that minimized the KL divergence between the participant's distribution (marginal shifted eg. on the range [-10,10]) and the reference distribution. The found value is set as the delay for such participant and could be either positive, negative or zero. A negative value means that the participant is slower than the participant's reference. A positive value means that the participant is faster than the reference. Zero means that the participant is as fast (or as slow) as the reference. However, only the negative values or zero are convenient for our modeling purposes because they could be treated as participant's delay as in the previous section. If so, the constraint can be solved either by applying a general shift to all of the resulting values in order to have only zero or negative values or by selecting another reference distribution (eg. the fastest participant: 's7'. This participant made in average the minimum number of fixations before to stop reading a paragraph) and recomputing again the Kullback-Leibler divergence to get new values. Table 4.8 shows the participant's delay values found by using the procedure described. From the table we can notice that most of the delay values are negatives (as it was expected) except for the reference (delay of 0). Two particular cases are 's9' and 's11' who were not considered for the analysis purposes because from a very basic statistical analysis they were set as the extreme participants: 's9' is the slowest participant (he read all the paragraph by doing many refixations before to make a decision) and 's11' the fastest participant (he read only a few words to make a decision). So, their behavior were considered as too different with most of the participant. Fig. 4.23 plots the participant's delay curve.

Finally, the participant's delays learned with this approach were also tested by running the model and we proceeded as before, by computing both empirical fixation distributions: the no-abandon and the abandon distributions in the $Cos \times Rank$ space. Both empirical distributions are shown in Figure 4.24. As before, a Bayesian classifier and a parametric approach to fit both distributions were also used. Equations were the same for the Bayesian classifier (Eq. 4.3, pag. 66), the two class-conditional probabilities density functions (Eqs. 4.4 and 4.5, pag. 66) and the decision rule (Eq. 4.6, pag. 66). In the case of the no-abandon distribution, Figure 4.25 (top, left) shows the empirical marginal distribution $\hat{p}_R(r|\overline{Ab})$. As the *Rank* increases, the probability of not abandoning the paragraph decreases. This evolution was also modeled with a sigmoid function (Eq. 4.7, pag. 66). Concerning the probability density function $p_{C|R}(.)$, the natural model (Fig. 4.24a) is also a Gamma one whose parameters depend on the *Rank* value. Then we have:

$$p_{C|R}(c|R = r, \overline{Ab}) = \frac{A(r)c^{k(r)-1}}{(k(r)-1)!\theta(r)^{k(r)}} e^{-\frac{c}{\theta(r)}}, p_R(r, \overline{Ab}) = \varphi(r)\,.$$

Participant	Delay
s1	-3
s2	-4
s3	-1
s4	-8
s5	-5
s6	-4
s7	0
s8	-3
s9	-10
s10	-3
s11	3
s13	-4
s14	-3
s17	-4
s18	-2
s19	-1
s20	-1
s21	-2

Table 4.8: Participant's delays learned by using the KL divergence. Participant 's7' was chosen as reference. Delays of participants 's9' and 's11' are not considered for our modeling purposes.

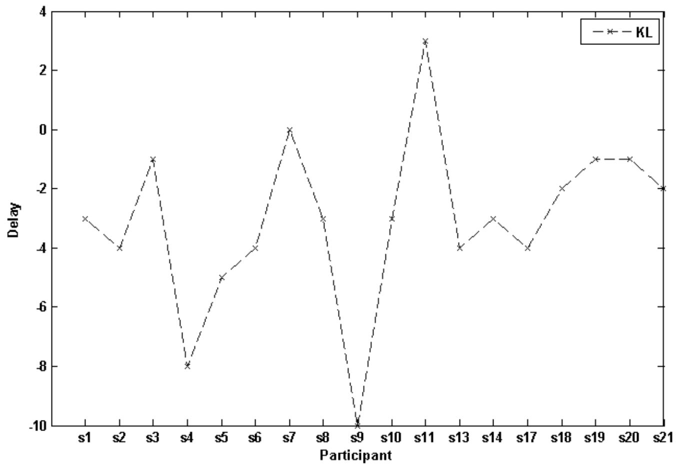
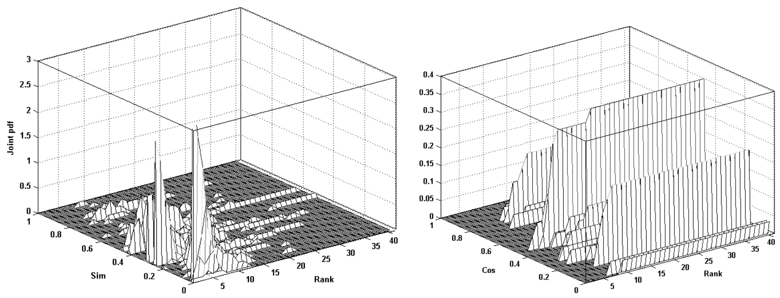


Figure 4.23: Participant's delays learned by using the KL divergence.



(a) (b)

Figure 4.24: a) Empirical "no-abandon" distribution $\hat{p}_{CR}(c, r|\overline{Ab})$ and b) "abandon" distribution $\hat{p}_{CR}(c, r|Ab)$ in the $Cos \times Rank$ space. Participant's delays learned by using the KL divergence were applied.

The shape $k(r)$ increases and the scale $\theta(r)$ decreases squarely (Fig. 4.25, left column). Both are modeled by squared equations: $k(r) = Tr^2+Ur+V$ and $\theta(r) = Wr^2+Xr+Z$. We then obtained eight independent parameters to model the complete "no-abandon" joint distribution: offset r_0 and slope α for the sigmoid $\varphi(r)$, the coefficients T,U,V for $k(r)$ and W,X,Z for $\theta(r)$.

In the case of the abandon distribution, the marginal pdf $\hat{p}_R(r|Ab)$ was modeled with another sigmoid function $\varphi'(r)$ (Fig. 4.25, top right). But here, it is an increasing function. The conditional distribution $\hat{p}_{C|R}(c|R = r, Ab)$ is a Gamma distribution with a shape $k'(r)$ and a scale $\theta'(r)$. The shape $k'(r)$ exponentially decreases while the scale $\theta'(r)$ exponentially increases (Fig. 4.25, right column). Both are modeled by exponential equations: $k'(r) = (s_2 - s_1)e^{-r/p_0} + s_2$ and $\theta'(r) = (t_2 - t_1)e^{-r/q_0} + t_2$. Equations of the pdf are the same as the previous case, but with a different set of functions $\{\varphi'(r), k'(r), \theta'(r)\}$ which gives us eight parameters (offset r_0 and slope α for the sigmoid $\varphi'(r)$; p_0,s_1 and s_2 for $k'(r)$; q_0, t_1 and t_2 for $\theta'(r)$).

The prior probabilities $P(Ab) = 0.69$ and $P(\overline{Ab}) = 0.31$ were learned from data as before. Then, the total number of learned parameters was then 17 (8+8+1). Table 4.9 summarizes the total number of parameters and their values for both modeled distributions: the no-abandon and the abandon.

The intersection between the two modeled distributions $P(\overline{Ab}|c, r)$ and $P(Ab|c, r)$ after learning is oblique as it is shown in Fig. 4.26. As before, the frontier does not seem to be a linear function as is shown in Fig. 4.27. The frontier was approximated by an exponential function in the $Cos \times Rank$ space: $Cos_0 = 99 \times \frac{1}{Rank^2} - 0.2$. However, most of the data are concentrated for Cos values below of 0.8. This situation can be seen with the superposition of densities of both posterior probabilities of Fig. 4.28a and 4.28b. For Cos values below of 0.8 the frontier seems to be linear as is shown in Fig. 4.27. Then, the frontier was also approximated by a linear function: $Cos_0 = -0.101 \times Rank + 1.727$.

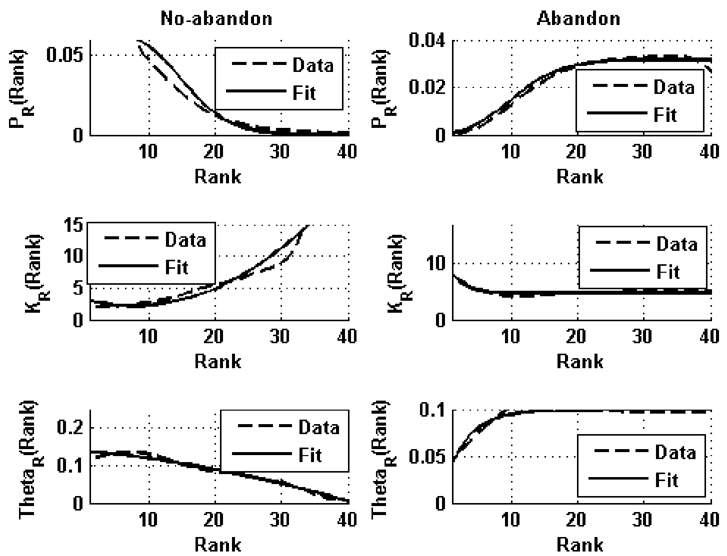


Figure 4.25: Data and fitting of marginal distributions, shape and scale for the "no-abandon" and "abandon" distributions. Participant's delays learned by using the KL divergence were applied.

Distribution	Parameters
No-abandon	r_0=15.0, α=0.292, T= 0.007, U=0.027, V=1.777 W= $-2.089e-004$, X= 0.002, Z= 0.124
Abandon	r'_0=10.0, α'= 0.299, p_0 = 1.759,s_1= 8.076, s_2=4.352, q_0= 3.359 t_1= 0.032, t_2=0.105

Table 4.9: Model parameters for both the no-abandon and the abandon distributions. Participant's delays learned by using the KL divergence were applied.

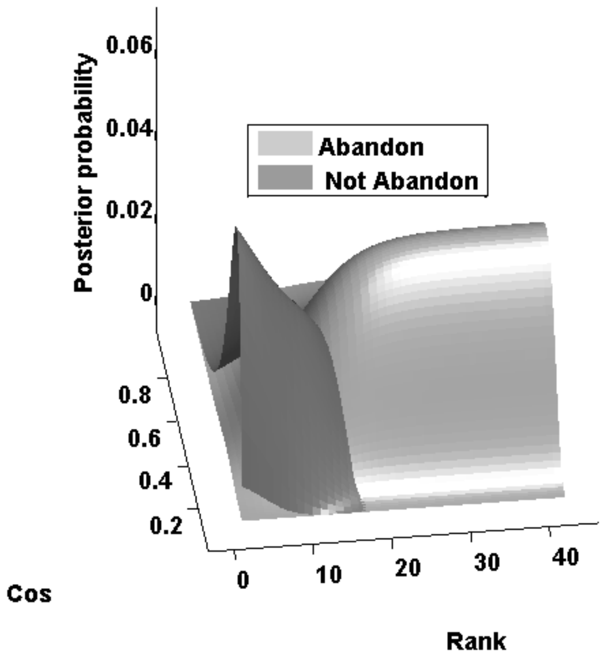


Figure 4.26: The posterior prob. $P(\overline{Ab}|c, r)$ and $P(Ab|c, r)$ in the $Cos \times Rank$ space. Participant's delays learned by using the KL divergence were applied.

At the end, the linear approximation (2 parameters) was set rather than the exponential approximation (3 parameters) due its lower complexity with respect to the number of parameters and because during the evaluation of the model performance, we got the same results for both approximations as it will be discussed now.

The model is ran it on one third of the data. The average difference between the ranks at which model and participant stopped reading was computed only for the cases where the model stop reading either before or at the same time of participant. This was observed in 59.38% of the cases and in 40.62% of the cases the model stopped reading after the participant. We got a value of 3.05 (SE=0.83) and 3.1 (SE=0.5) for the linear and exponential approximations respectively. Then, the linear approximation was preferred due its lower complexity. A value of 3.05 (SE=0.83) means that, in the average, the model stops reading 3.05 fixations earlier than participants. The distribution of the differences between the number of fixations made by the model and the participant is showed in Figure 4.29. The difference between the rank at which model and participant stopped reading is indicated on the horizontal axis and the number of occurrences on the vertical axis. A negative value means that our model is more reactive than participant and viceversa. Zero means that the model is resulting as reactive as participant. A total of 32 cases (scanpaths) were considered in the testing set.

For comparison purposes with the random model built (Ch. 3, pag. 55), the best

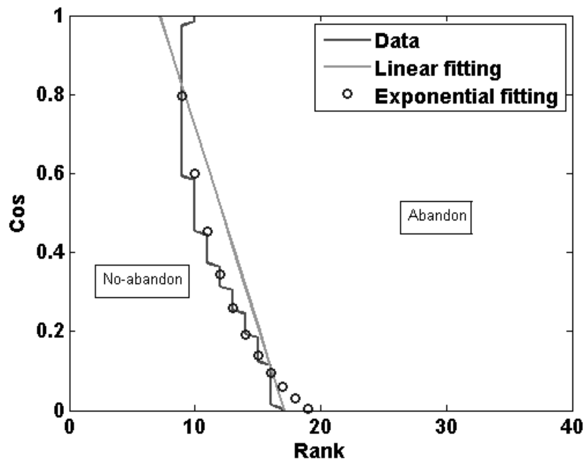
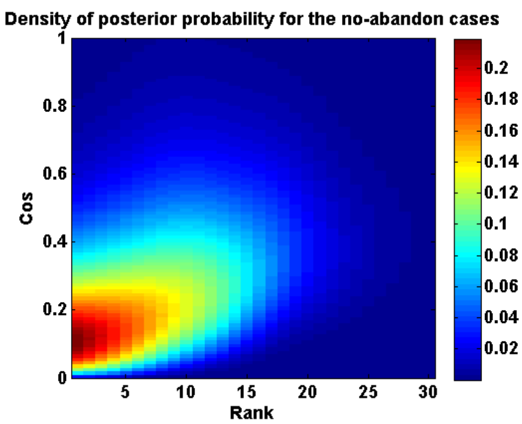
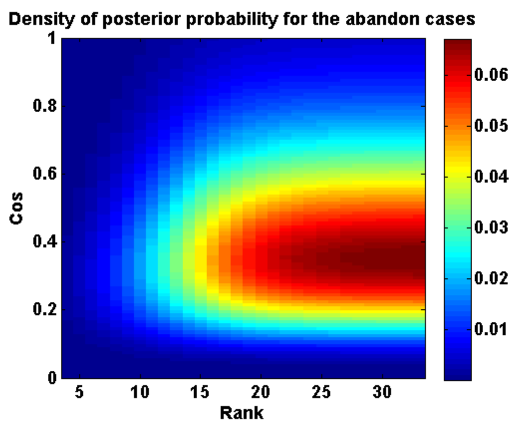


Figure 4.27: The Bayesian frontier between both the abandon and the no-abandon cases. Participant's delays learned by using the KL divergence were applied. Experimental data, their linear and exponential fitting are shown.



(a) No-Abandon



(b) Abandon

Figure 4.28: Densities of posterior probabilities for the no-abandon and the abandon distributions. Participant's delays learned by using the KL divergence were applied.

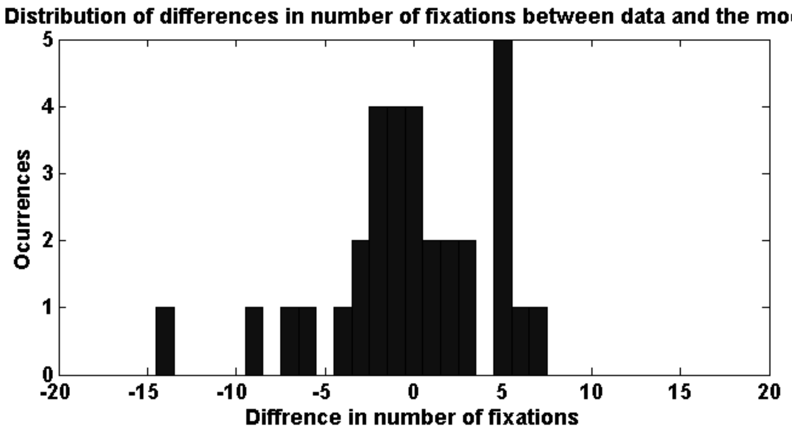


Figure 4.29: Distribution of the differences between the number of fixations (ranks) made by the participant and the model when. Delays learned by using the KL divergence were applied.

p value found such as the random model had the smallest average difference with participant ranks was for $p = 0.18$. That smallest average difference was of 9.43 (SE=1.2). Our model therefore appears to be much better than the best random model. More still important is that the model performance is better than the first model presented in about 1.7 fixations (3.05 fix., SE=0.83 vs 4.7 fix, SE=0.72) in this chapter when any consideration of a participant's delay was considered.

4.7 Conclusions

So far in the current chapter, we have presented a first version of our computational cognitive model to simulate people's behavior during a task of information search. More specifically, our model makes predictions at the level of words that are likely to be fixated by people before to stop reading a paragraph. Then, we have followed several approaches by trying to improve the performance of this first model. In all the approaches, the idea was to get a signature d for each participant. Our hypothesis was that participants do not make their decisions until the last fixation of a scanpath but they make it as is indicated by the delay value at fixation $(n - d)$.

The first approaches were based on special words: *step*, *ramp* and *target*. However, only in the case of the *ramp* word, we got a result but as we have argued before, this should be taken with caution.

In our last approach, finally we ended up with a second version of the model. Here, each participant was characterised with a delay value learned by using the KL divergence. And as we have noticed earlier, the model performance is better than the first version of our model. However, we thought that the idea to get a participant's signature (as a delay value) has not been exploited completely and there is at least one reason to think about that. For example, we have tried to characterise to participants with a simple value. Once this value is learned, the value is set and it is used for all the scan-paths of such participant. Maybe we should consider that the value is also modulated by another factor like the semantic association between the processed words so far and the search goal. Saying differently, a participant might not need to do the same number of fixations to confirm his decision if the semantic association between the words seen and the goal is of 0.4 than if the semantic association is 0.8 or higher. By this way, the participant's delay should be modulated by the semantic association between the words actually processed in the paragraph and the search goal during the reading process.

The evaluation of the model performance was done by following the experimental scanpaths. A new Cos_0 value is computed in each fixation while it is moving forward in a paragraph. However, if we had a new paragraph, our model will assume a linear exploration of words, although we know that this is not exactly the case in information search (Chanceaux et al., 2009).

Chapter 5

Computational Cognitive Modeling: a second environment

In the previous chapter, we have presented a computational cognitive model to simulate people's behavior when they are faced to a simple information search task. In this chapter we present a new computational cognitive model for another common task performed by people during information search: reading a paragraph with the task of quickly deciding whether it is better related to a given goal than another paragraph processed previously. Saying differently, we are interested in studying people's behavior when they are concerned with deciding whether the current paragraph is *more* interesting or not than another paragraph that has been processed before. The model corresponds to the Situation 2 (p. 40, Fig. 3.2) that we aim at modeling.

The goal was therefore to build a model that would describe the cognitive processes involved in this decision-making situation but also predict when an average user is likely to stop reading the current paragraph for a given goal and a paragraph processed previously. For the conception of the model, we have followed the same eye-tracking experimental approach to gather some data as it was done in last chapter. On next section, the new experiment is presented. Then, the learning phase of the model and the model performance are also discussed.

5.1 Experiment II: two-paragraph experiment

The experiment was intended to emphasize the decision to stop reading a paragraph with the task to decide if it is better related to a search goal than another paragraph processed previously.

5.1.1 Participants

A total of 22 French native speakers participated in the experiment (age range 22-56 years, mean age=29.3 years, sd=8.3 years). All participants had normal or corrected-to-normal visual acuity. They were naive with respect to the purpose of the study, and all gave written consent.

5.1.2 Material

The material described in Chapter 3 (p. 47) was used in the experiment: twenty goals with seven paragraphs each: two highly related (HR), two moderately related (MR) and three unrelated (UR) paragraphs.

5.1.3 Apparatus

The description of the eye-tracker device is given in Chapter 3 (p. 48). The EyeLink system was used in tracking mode sampling at 500Hz. Eye position data were collected only for the guiding eye.

5.1.4 Procedure

The experiment is composed of 20 trials, each one corresponding to a goal, in random order. During each trial participants were exposed to six stimuli pages composed of two paragraphs of the corresponding goal. At the very beginning of the experiment, an initial 9-point calibration was done. Drift corrections were done before each trial to ensure the recording accuracy of eye movements. The procedure began with a learning trial using one new goal and its corresponding seven paragraphs, excluded from the analyses. In each trial, a fixation cross is displayed at the top (horizontally centered) waiting for the participant's gaze to be inside a box of 120 pixels wide and 180 pixels high centered on the fixation cross. Once the participant's gaze is detected inside the box, the gaze position was controlled to remain inside during more than a fixed value of 500 ms. A re-calibration routine was automatically done if the timeout on the initial fixation cross elapsed or if the experimenter decided to run it. Then two paragraphs were presented together to the participant, as well as the goal as is shown in Fig. 5.1. Afterwards, the participant should select which paragraph is best related to the goal, by typing one key. The chosen paragraph is kept and the other is replaced by a new one. The participant should again select the most related to the goal. Then another paragraph replaces the one that was not selected and so on. This procedure is repeated until all seven paragraphs of the current goal were displayed. Participants rated their confidence in their selection. The participant knew that between each goal, a pause could be managed if necessary and he/she knew the number of remaining goals until the end. Figure 5.2 shows the different screens presented to the participant during a complete trial. In order to get a more rich process of competition along the trial a combination of paragraphs of the same type is avoided only for the first stimuli page.

During the experiment, participant's head position was stabilized with a chin rest and their forehead supported on a fixed bar. Participants were seated 57 cm in front of a 22-inch monitor (39°x 29° of visual field) with a screen resolution of 1024 by 768 pixels. The text was displayed at the centre of the screen (27°x 18° of visual field).

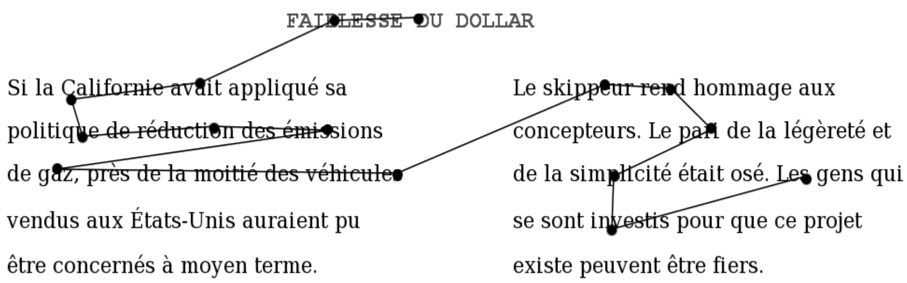


Figure 5.1: Example of material and scanpath.

In average, the text is displayed with 64 characters per line, corresponding about 5 characters in fovea. We wrote our experiments in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

5.2 Pre-processing data

All the procedure followed is detailed in Chapter 3 (p. 48). Basically it includes the conversion from eye-tracker data to a file containing the words actually processed in each fixation by using our window-based approach.

Once data were adequate, and before to do a deeply analysis, some considerations and the computation of some basic statistics were done. Their descriptions are now given.

5.3 Basic statistics

Due to the nature of our experiment that has the aim to emphasize the participant's decision to selecting one of two paragraphs presented according to its relatedness to the goal, we thought that it was interesting to only consider and to study the following cases or participant's behaviors:

(a) Participant has only to collect information. Participant is only collecting information when he/she is discovering a new paragraph and he/she does not know anything about other paragraph. This is the case when the first stimuli page is displayed and participant is reading the first paragraph.

(b) Participant has to read a paragraph to compare it to the other one. When participant decides to leave the first paragraph visited and goes to the other one, he is performing two activities at the same time: collecting information and comparing the current paragraph with the previous in order to choose the best one.

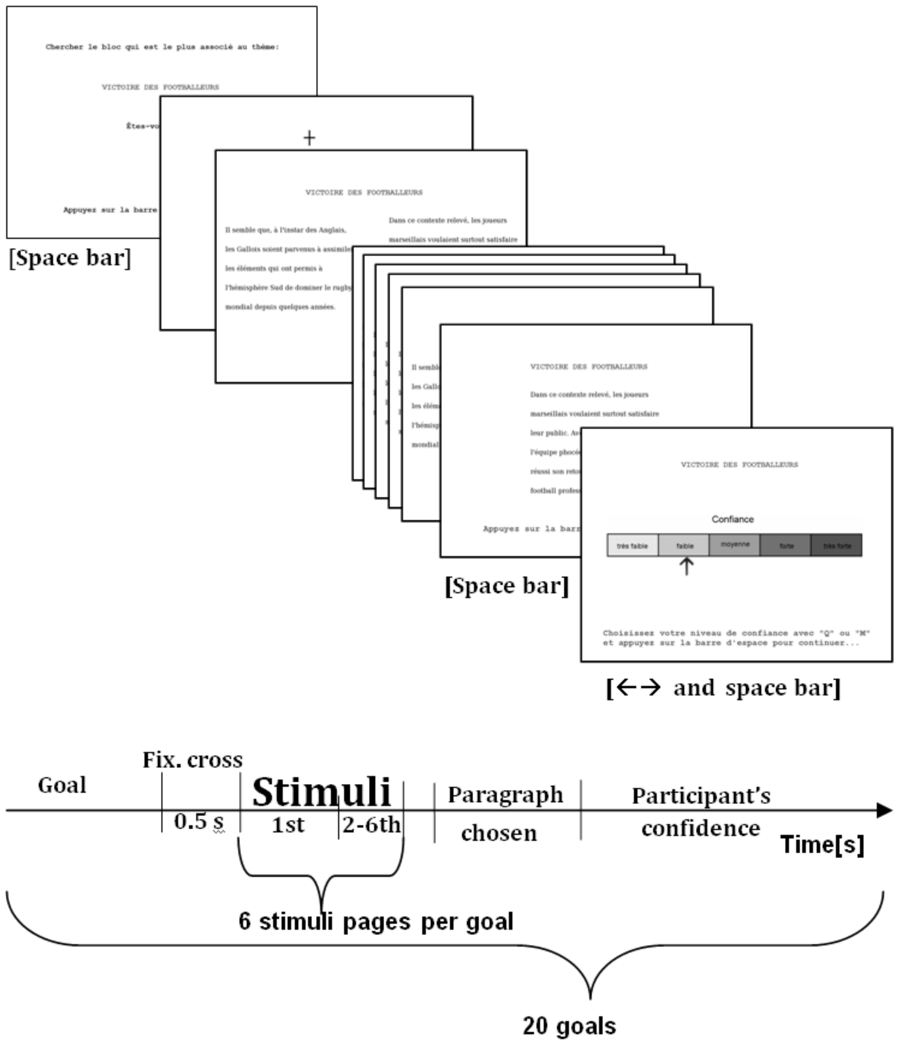


Figure 5.2: Temporal sequence of a trial composed of the different screens of Experiment II.

In order to study these two behaviors, different categories of visiting a paragraph were considered. A visit to a paragraph of type HR or MR is referred to simply as HM. When people are visiting a paragraph HM, we make a distinction between three cases:

- People does not know anything about the other paragraph. This case is notated as HM/0 and can be read as *visiting a paragraph HM without knowing anything about the other one*. This case corresponds to the first behavior (a) of our interest.
- People knows the other paragraph and this paragraph is also of type HM. This case is notated as HM/HM and can be read as *visiting a paragraph HM knowing the previous paragraph of type HM*. This case corresponds to the behavior (b) mentioned above.
- People knows the other paragraph and this paragraph is of type UR. This case is notated as HM/UR and can be read as *visiting a paragraph HM knowing the previous paragraph of type UR*. This case also corresponds to the second behavior (b) already mentioned.

To sum up, we analyzed the three categories of visiting a paragraph: HM/0, HM/HM and HM/UR. Although, paragraphs HR and MR have different association with the goal, both are considered into the same category of visiting a paragraph (HM) in order to reduce the number of categories treated but paragraphs of type UR must be considered in a separated category since they are unrelated to the goal. In such categories, we have analyzed the number of fixations per word, fixation duration and the shape of the scanpath or "strategy" that people follow when they are visiting the paragraphs.

Number of fixations per word. This indicator can be seen as a reading rate. The number of fixations in a paragraph is divided by the total number of words in the paragraph. Figure 5.3 shows the results for the visits HM/0, HM/UR and HM/HM. There is not a significant difference of number of fixations between visits of type HM/0 and HM/UR (t-test, $p>0.96$), but there is a significant difference in between HM/0 and HM/HM (t-test, $p<0.05$) and HM/UR and HM/HM (t-test, $p<0.05$). It means that reading the first paragraph or reading one when the previous one is unrelated to the goal is about the same with respect to the number of fixations per word. Thus, visits of type HM/0 and HM/UR can be considered together and called: HM/0-HM/UR. With this consideration, results show that the number of fixations in visits of type HM/HM (mean=0.68, SD=0.10) is greater than in visits of type HM/0-HM/UR (mean=0.36, SD=0.06) and significantly different (t-test, $p<0.05$). One of the reasons may be due to the fact that when participant is visiting a paragraph of type HM knowing the other paragraph of type HM, it is not easy to decide which one is the best because both paragraphs have an association with the goal, so participant might revisit the paragraphs in order to confirm their knowledge about them and being able to choose one. Contrarily,

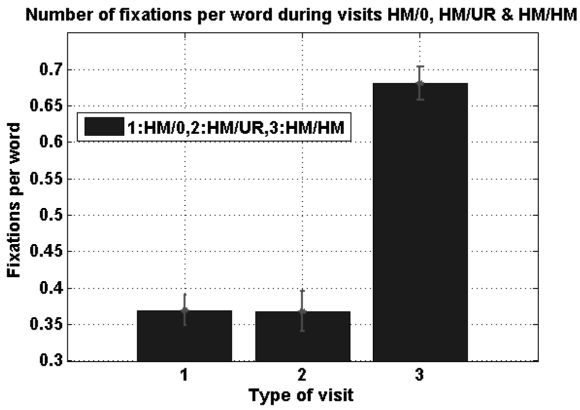


Figure 5.3: Number of fixations per word during visits: HM/0, HM/UR and HM/HM.

in the case of visits of type HM/0-HM/UR the participant is only engaged in a task of collecting information even in the case of HM/UR, since paragraphs of type UR do not have any association with the goal and participants do not need to do a lot of fixations.

Fixation duration. Secondly, the fixation duration is analyzed. Figure 5.4 shows the mean duration of fixations during visits of type HM/0, HM/UR and HM/HM. Results shows that there is not a significant difference on fixation duration during visits of type HM/UR and HM/0 (t-test, $p>0.82$) as in the previous case, but there is a significant difference in between HM/0 and HM/HM (t-test, $p<0.05$) and HM/UR and HM/HM (t-test, $p<0.05$). By regrouping as previously HM/0 and HM/UR, the duration on fixations in visits of type HM/HM (mean=176.90 ms, SD=5.13) is longer than in visits of type HM/UR-HM/0 (mean=171.45 ms, SD=5.57) and significantly different (t-test, $p<0.05$). The reason may be due to the fact that when participant is visiting a paragraph of type HM knowing the other paragraph of type HM, making a decision is not easy because both paragraphs have a strong association with the goal and participant may spend more time in each fixation in order to be able to choose the best. In the case of visits of type HM/0-HM/UR, participants are only collecting information and they do not need to spend a lot of time on each fixation.

"Strategy" for visiting paragraphs. Our interest here is to know whether the previous results still apply with respect to the visual strategies used by the participants while visiting a paragraph. Fixations are considered by pairs to determine the visual strategy for visiting a paragraph: the current fixation and the previous one. The shape of the ingoing saccade from the previous to the current fixation is classified as *scanning* or *reading*. Our approach is not based on the reading rate as in the case of Carver (1990). Fig. 5.5 shows an example of a scanpath with 9 fixations. Fixations are indicated by

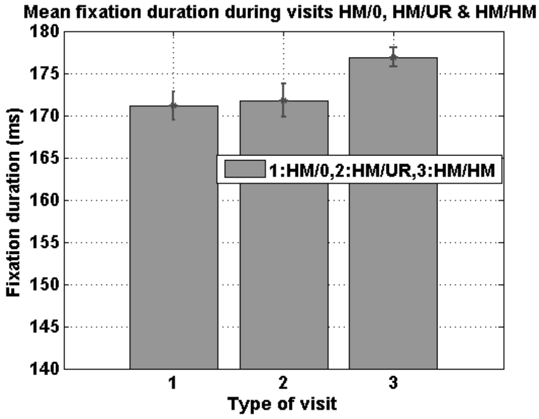


Figure 5.4: Mean fixation duration during visits: HM/0, HM/UR and HM/HM.

the gray spots whereas the saccades by the lines connecting pairs of fixations. Fixations are labeled as "R" (*reading*) or "S" (*scanning*) according to the shape of the saccade that connect them. For the case of *reading*, basically we consider either a short or long forward saccade (a maximum of three words length) on the same line of text, a short backward saccade (a maximum of one word length) or a long forward saccade going from the last word of the current line to the first word of the next line of the paragraph. In any other case, the scanpath is classified as *scanning*. The first fixation of a scanpath is always labeled as *scanning* ("S"). In the example of the Fig. 5.5, fixation 1 is labeled as "S". Then there is a long forward saccade to the fixation 2 ("R"), a short backward saccade to the fixation 3 ("R"). Then there are two long forward saccades to the fixation 4 ("R") and 5 ("R"). Fixation 6 is labeled as "R" because there is long forward saccade going from the fixation 5 to the fixation 6. Ingoing saccades from fixations 7 and 8 are classified as *scanning* ("S"). At the end, there is short forward saccade to the fixation 9 ("R").

Once the saccades were classified, we set a value of 0 for a *reading* saccade and 1 for a *scanning* saccade. Then, we computed the average of all these values to have an idea if participant's trend is more likely to perform a *reading* or a *scanning* strategy. A value close to 0 means a linear *reading* and a value close to 1 means a *scanning* strategy or non-linear reading. Figure 5.6 shows the strategy for visiting a paragraph during visits of type HM/0, HM/UR and HM/HM. Contrarily to the number of fixations per word and fixation duration, there is a significant difference between visits of type HM/0 and HM/UR (t-test, $p<0.05$). Also, there is a significant difference between HM/HM and HM/UR (t-test: $p<0.05$) but there is not a significant difference between HM/HM and HM/0 (t-test: $p>0.20$).

As expected, in no case participants performed either a pure *scanning* or a pure *reading* strategy. However, during visits of type HM/UR (mean=0.42, SD=0.06),

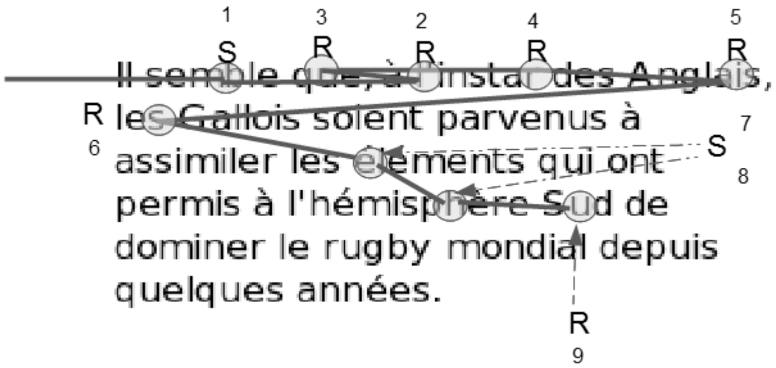


Figure 5.5: Strategy for visiting a paragraph for HM/0, HM/UR and HM/HM visits.

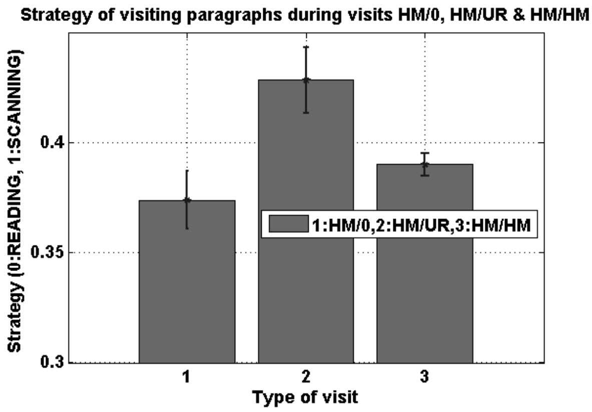


Figure 5.6: Strategy for visiting a paragraph for HM/0, HM/UR and HM/HM visits.

participant's trend is to *scan* more than *read* as opposed in the case of visits HM/0 (mean=0.37, SD=0.05) and HM/HM (mean=0.39, SD=0.02). This situation may be due to the fact that when people are looking at the current paragraph HM but they have realized that the other paragraph is irrelevant to the goal (paragraph UR), they are encouraged to find quickly if the current paragraph is related or not to the goal. Therefore, many words are not fixated. In the case of visits HM/HM, making a decision is not easy because both paragraphs are related to the goal, so participant should read carefully their contents before being able to choose the best

Preference over first or second paragraph seen. As we have described before, in each trial participants were exposed to two paragraphs HR (here namely hr_a and hr_b), two paragraphs MR and three paragraphs UR. Due to the fact that hr_a and hr_b have strong association with the goal, they are much more likely to be chosen as the best one at the end of the trial. However, one of them is displayed before the other one, and here, we investigated how people make a decision in between them: is there a preference to choose the first paragraph versus the second paragraph seen? In other words, our question is, do people tend to maintain their initial choice or not? We studied such people's behavior by computing the probability of choosing one of the two paragraphs. The probability that participants choose the first paragraph hr_1 over the second paragraph hr_2 seen $P(Hr^* = hr_1)$ is computed as:

$$P(Hr_1 = hr_a) \times P(Hr^* = hr_1/hr_1 = hr_a) + P(Hr_1 = hr_b) \times P(Hr^* = hr_1/hr_1 = hr_b) \tag{5.1}$$

in which Hr^* is the paragraph chosen and $P(Hr^* = hr_2) = 1 - P(Hr^* = hr_1)$.

Results showed that there is not an important participant's preference to choose the first paragraph seen over the second one: $P(Hr^* = hr_1)$, mean=0.53, SD=0.18; $P(Hr^* = hr_1)$, mean=0.46, SD=0.18 (t-test: p>0.36).

Preference over left or right side. In a similar way, the preference over the paragraph displayed on the left or on the right side was investigated. The probability that participants choose the left paragraph (namely here hr_{left}) over the right one (hr_{right}) is computed using the Eq. 5.1 by replacing hr_1 for hr_{left} and hr_2 for hr_{right}.

Our results show that there is not an important effect (t-test: p>0.14) on participant's decision of whether a paragraph is displayed on left (mean=0.55, SD=0.14) or right (mean=0.45, SD=0.14) side.

Finally, the last aspect reviewed before to move to the modeling work is detailed now.

Gaze cascade effect. Shimojo et al. (2003) tested how subjects orient their attention in both preference and non-preference tasks. Their main result is that, in choices between two alternative forced-choice tasks, subjects exhibit a statistically significant tendency to gaze increasingly towards the chosen object. It supports the hypothesis

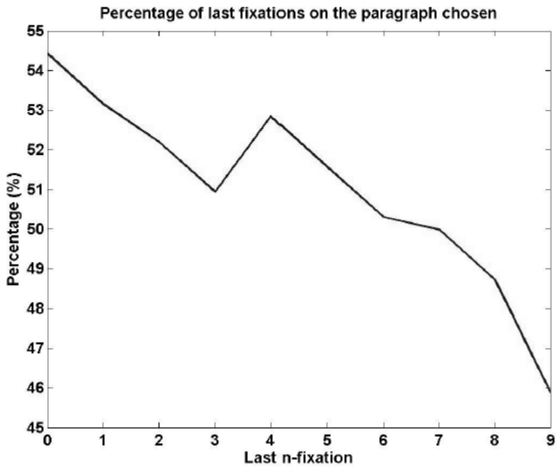


Figure 5.7: Percentage of occurrences of the last 10 fixations on the chosen paragraph.

that gaze participates directly in the preference formation process by involving the gradual gaze shift toward the item that is eventually chosen as more “attractive”. The phenomenon is called “Gaze cascade effect”.

And here, we investigated in a very roughly way if this effect occurs in our experiment. To do that, we have analyzed the last 10 fixations of all scanpaths involving visits to paragraphs of type HR. We only considered the last fixations but we also could have considered a short period of time (eg. the last 5 seconds of a trial for instance or similar). Our interest was to look if the gaze is shifted toward the paragraph that is chosen as the best at the end and the final decision might be taken during this set of fixations.

Our data shows that the gaze cascade effect seems to occur in such a decision task on textual material as is illustrated in Fig. 5.7. It shows the percentage of occurrences when the last 10 fixations $n, n-1, ..., n-9$ were on the paragraph chosen at the end. n is the last fixation. There is an increased tendency from fixation $n-9$ to n of the percentage of fixations (46% at fixation $n-9$ and around 54% at fixation n) over such paragraph.

5.4 Modeling

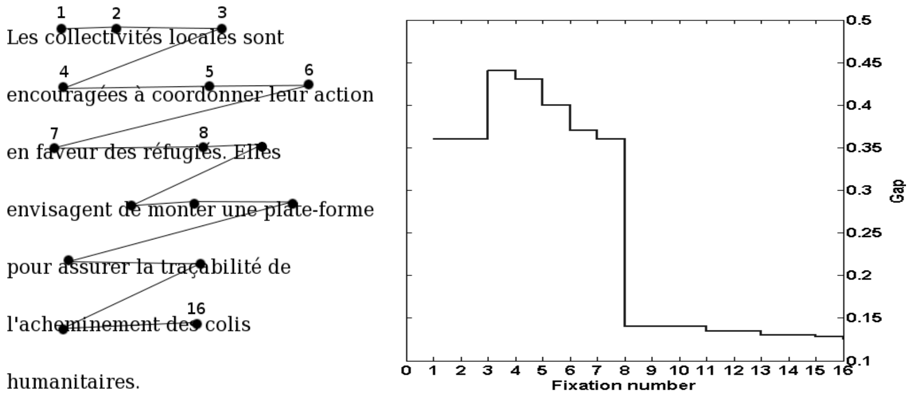
Making predictions in processing a paragraph. The model should be able to predict the way a paragraph is processed given a paragraph already processed and a goal. For example, given the left paragraph of Fig. 5.1 and the goal, the model should be able to predict the way an average user would process the right paragraph. In this example, the paragraph is processed partially and abandoned after about a dozen words have been fixated. Paragraphs can be examined several times by participants during a trial, but we restricted our analysis to first visits of the current paragraph. It is also worth noting that the previous paragraph is not necessary on the same stimuli page as the current paragraph. It could have been seen on the previous stimuli page. That is for instance the case of the left paragraph of Fig. 5.1 which has been processed with another paragraph in mind, seen on the previous stimuli page.

Modeling Semantic Judgments. As in the previous modeling work, LSA (Landauer and Dutnais, 1997; Landauer et al., 2007) is also used to dynamically compute the semantic similarities between the goal and each set of words that are supposed to has been fixated.

Effect of the Prior Paragraph. We suspect that the relatedness of the prior paragraph to the goal may play a role in the way the current paragraph is processed as we have shown with the basic statistics of the previous section. In order to simplify the design of the model, we therefore analyzed two extreme cases: the words actually processed in the prior paragraph are strongly related to the goal (HR) or they are not related at all to the goal (UR). We used two thresholds of cosine similarity for that, which were set to 0.05 and 0.25. Paragraphs whose semantic similarity with the goal falls in between were not considered. The first case is called C|HR (read the Current knowing that the previous one is Highly Related) and the second one is called C|UR (Current | Previous=Unrelated). We also analyzed cases when no prior paragraph exists, called C|0 (Current | Nothing). The basic statistics already presented show that in terms of number of fixations, and fixation duration, C|UR=C|0 and both are significantly different from C|HR. It means that reading a paragraph while the other one is not related to the goal is similar to reading the very first paragraph, without information about a prior paragraph. Therefore we will only consider the case C|HR in this modeling work: reading a paragraph with another one in mind which is highly related to the goal.

5.4.1 Modeling the Decision

Two variables involved. We investigated which variables could played a role in the decision to stop reading a paragraph with the task of quickly deciding whether it is better related to a given goal than another paragraph processed previously. As it was mentioned in Chapter 3 (p. 52), the semantic similarity between the words processed in such paragraphs and the goal must be involved in such decision. Then, we suspected that the decision is made when the difference between the current (*cp*) and the previous paragraph (*pp*) is large enough to know for sure which one is the best. If they are too close to each other, no decision can be made and reading is pursued. The association



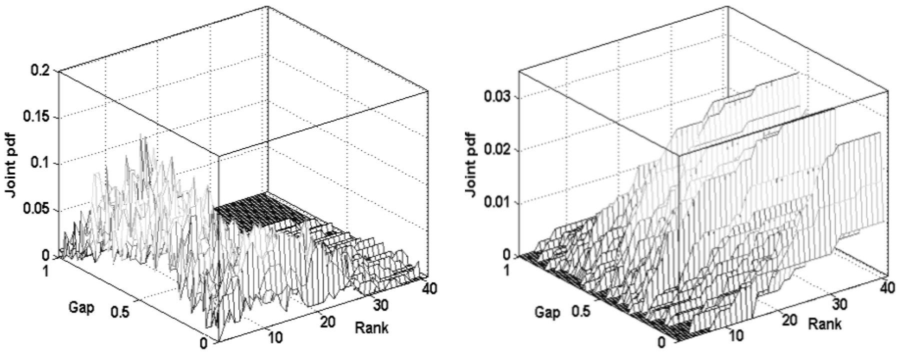
(a) (b)

Figure 5.8: a) Example of scanpath in the C|HR condition. b) Its *Gap* evolution.

to the goal g is obviously involved in that perception of a difference between the two paragraphs. In the Experiment 1, the relatedness with the goal played a role, but here there are two cosine similarities and it is likely that their difference is important. For that reason, we defined a variable called *Gap* as follows:

$$Gap = |sim(\text{words processed of } pp,\, g) - sim(\text{words processed of } cp,\, g)|\,. \qquad (5.2)$$

in which *sim* is the LSA cosine between the two vectors of words actually processed in paragraphs *pp* and *cp* with the goal g. *Gap* changes constantly while a paragraph is processed. When the two paragraphs are equally similar to the goal, that variable is zero. When one paragraph is much more associated to the goal than the other one, that variable has a high value. *Gap* can be easily calculated dynamically, after each word of the current paragraph has been processed. Consider for example Fig. 5.8a. Suppose that a prior paragraph has already been visited (paragraph and goal are not shown) and the sequence of words processed so far has led to a similarity sim_1 with the goal *"associations humanitaires"* of 0.62. In the first two fixations on the current paragraph, only the word *"collectivités"* is supposed to have been processed according to our window-based prediction. Therefore in both cases $Gap = |sim_1 - sim($*"collectivités"*,*"associations humanitaires"*$)|$ = 0.62 - 0.26 = 0.36. During fixation 3, two extra words were processed leading to a new value of $Gap = |sim_1 - sim($*"collectivités locales sont"*,*"associations humanitaires"*$)| = 0.44$. In fixation 4, $Gap = |sim_1 - sim($*"collectivités locales sont encouragées à"*, *"associations humanitaires"*$)| = 0.43$. In fixation number 5, $Gap = |sim_1 - sim($*"collectivités locales sont encouragées à coordonner leur"*, *"associations humanitaires"*$)| = 0.40$. In fixation 8, the *Gap* value dropped to 0.14 because of the word *"réfugiés"* which makes the LSA vector much more similar to the goal vector. Figure 5.8b shows the evolution of the *Gap* value along



(a) (b)

Figure 5.9: a) Empirical “no-abandon” distribution $\hat{p}_{GR}(g, r|\overline{Ab})$ and b) “abandon” distribution $\hat{p}_{GR}(g, r|Ab)$ in the $Gap \times Rank$ space.

the fixations in the scanpath. This example illustrates that a high value of *Gap* may not directly induce the decision, in particular if it appears too early in the scanpath. We assume that the decision also depends on the number of words processed so far in the current paragraph exactly like in the Experiment 1. The more words processed, the higher the confidence in the perception of the difference between paragraphs. If only two or three words have been processed, it is less likely that *Gap* is accurate. Therefore, we assume that there should be a relationship between *Gap* and the number of words processed. As in our model for the Experiment I, the variable *Rank of fixation* was used as indicator of the number of words processed so far in the paragraph.

Once we have determined the two variables that could played the role in the decision to stop reading a paragraph, we followed the same approach used on the first modeling work in order to study how the decision depends on these variables. Two empirical fixation distributions were computed. Their descriptions are given now.

Abandon and no-abandon distributions. The distribution of the no-abandon cases and the distribution of the abandon cases were computed in order to learn the frontier between both cases, but now in the $Gap \times Rank$ space of participant data. Each participant fixation was associated to a point in the $Gap \times Rank$ space. *Rank* is also a discrete measure between 1 and the maximum number of fixations in the data. *Gap* has been computed according to Formula 5.2, taking into account the words already processed in each paragraph as well as the goal and discretized into one of 100 bins, from 0 to 1. The no-abandon distribution was computed by simply counting the number of fixations that did not lead to an abandon for each cell of the $Gap \times Rank$ grid. It concerns all fixations except the last one of each scanpath. The abandon distribution was built from all very last fixations of all scanpaths, including also the virtual fixations as it was done on

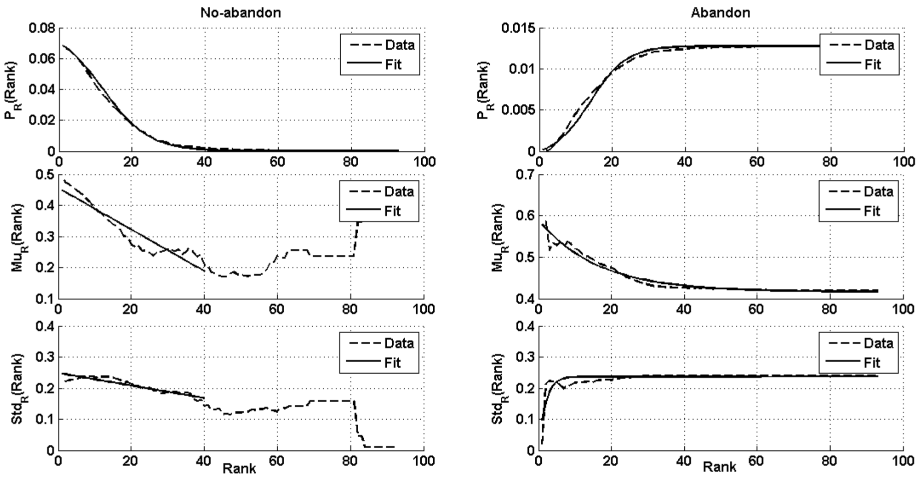


Figure 5.10: Data and fitting of marginal distribution, mean and standard deviation for the "no-abandon" and "abandon" distributions.

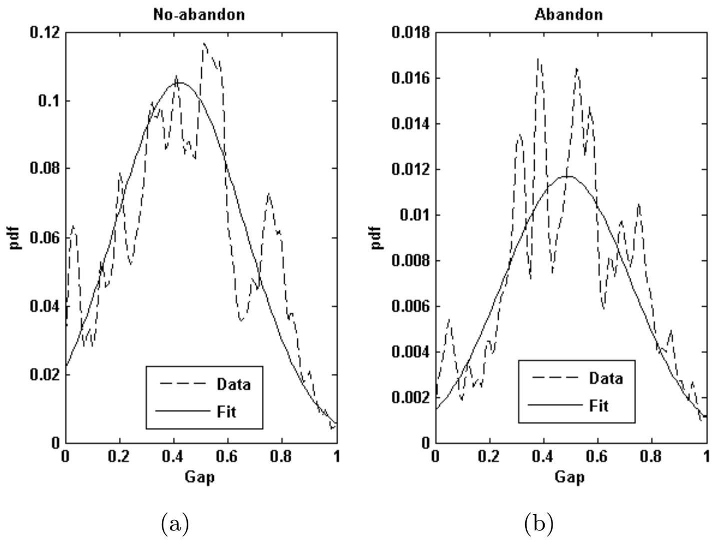


Figure 5.11: Example of empirical conditional pdf and fit with a gaussian model for *Rank* value at 10 for the a) "no-abandon" and b) "abandon" class.

Chapter 4 (p. 63), but here the *Gap* is considered instead of a *Cos* value. By using the frontier curve between these two behaviors (continue or stop reading) it is possible to predict the reader's behavior depending of the location of any observation (g,r) above or under the curve. To find the frontier, the methodology based on a Bayesian classifier is also used and outlined by Equations 3.3, 3.4, 3.5, 3.6, 3.7 and 3.8 (Ch. 3, p. 54). Now by replacing *Sim* with the corresponding variable *Gap* of Eq. 5.2 in the Equations 3.5, 3.6, 3.7 and 3.8, we have the Bayesian classifier (Eq. 5.3), the two class-conditional probabilities density functions (Eqs. 5.4 and 5.5) and the decision rule (Eq. 5.6) as follows:

$$P(\overline{Ab}) \times p_{GR}(g,r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{GR}(g,r|Ab)\,. \tag{5.3}$$

$$p_{GR}(g,r|\overline{Ab}) = p_{G|R}(g|R=r,\overline{Ab}) \times p_R(r|\overline{Ab}) \tag{5.4}$$

and

$$p_{GR}(g,r|Ab) = p_{G|R}(g|R=r,Ab) \times p_R(r|Ab) \tag{5.5}$$

$$P(\overline{Ab}) \times p_{G|R}(g|R=r,\overline{Ab}) \times p_R(r|\overline{Ab}) \underset{Ab}{\overset{\overline{Ab}}{\gtrless}} P(Ab) \times p_{G|R}(g|R=r,Ab) \times p_R(r|Ab) \tag{5.6}$$

Now the statistical parametric approach is explained in order to estimate the density functions and the prior probabilities.

Parametric model for the "no-abandon" distribution. Figure 5.10 (top, left) shows the empirical marginal distribution $\hat{p}_R(r|\overline{Ab})$. As the *Rank* increases, the probability of not abandoning the paragraph decreases. This evolution was modeled with a sigmoid function:

$$\varphi(r) = \frac{P_{RMax} \times (1+e^{-\alpha r_0})}{1+e^{\alpha(r-r_0)}}\,.$$

As before, there are only two parameters r_0 and α to fit because the integral is 1. With respect to the probability density function $p_{G|R}(.)$, the natural model (Fig. 5.9a) is a Gaussian one whose parameters depend on the *Rank* value. Figure 5.11a illustrates on a particular *Rank* the reason why the Gaussian is a good function to model these data. The mean $\mu(r)$ and the standard deviation $\sigma(r)$ linearly decrease (Fig. 5.10, left column). Both are modeled by simple linear equations: $\mu(r) = Mr + N$ and $\sigma(r) = Or + P$. The linear regressions are only performed up to the *Rank*=40 since that $\hat{p}_R(r > 40|\overline{Ab})$ is close to zero and there is no more enough data. Then we have:

$$p_{G|R}(g|R=r,\overline{Ab}) = \frac{A(r)}{\sqrt{2\pi}\sigma(r)} e^{-\frac{(g-\mu(r))^2}{2\sigma(r)^2}}\,, p_R(r,\overline{Ab}) = \varphi(r)\,.$$

Due to the *Gap* value is between 0 and 1, $A(r)$ is a normalization function to ensure that $p_{G|R}(g|R=r,\overline{Ab})$ is a probability density function: $A(r) = F_{\mu,\sigma}(1) - F_{\mu,\sigma}(0)$, with

Distribution	Parameters
No-abandon	r_0=12.0, α=0.16, M=-0.007, N=0.46, O=-0.0020, P=0.2479
Abandon	r'_0=14.0, α'=0.2, $p_0 = 16.38, s_1$=0.42, s_2=0.59, t_{max}=0.24, q_0=1.91

Table 5.1: Model parameters for both the no-abandon and the abandon distributions.

$F_{\mu,\sigma}(.)$ being the repartition function of a Gaussian distribution with a mean μ and a standard deviation σ. We then obtained six independent parameters to model the complete "no-abandon" joint distribution: offset r_0 and slope α for $\varphi(r)$, the coefficients M,N for $\mu(r)$ and O,P for $\sigma(r)$.

Parametric model for the "abandon" distribution. The marginal pdf $\hat{p}_R(r|Ab)$ was modeled with another sigmoid function $\varphi'(r)$ (Fig. 5.10, top right). But here, it is an increasing function. At *Rank* 0, there is no abandon and at the maximal *Rank* value, all scanpaths have shown an abandon. The conditional distribution $\hat{p}_{G|R}(g|R = r, Ab)$ is a Gaussian distribution with a mean $\mu'(r)$ and a standard deviation $\sigma'(r)$. As in the previous case, Fig. 5.11b illustrates how these data are modeling by a Gaussian function. The mean $\mu'(r)$ exponentially decreases while the standard deviation $\sigma'(r)$ exponentially increases (Fig. 5.10, right column). Both are modeled by exponential equations: $\mu'(r) = (s_2 - s_1)e^{-r/p_0} + s_2$ and $\sigma'(r) = t_{max}(1 - e^{-r/q_0})$. Equations of the pdf are the same as the previous case, but with a different set of functions $\{\varphi'(r),\mu'(r),\sigma'(r)\}$ which gives us seven parameters (r'_0 and α' for $\varphi'(r)$; p_0, s_1 and s_2 for $\mu'(r)$; t_{max} and q_0 for $\sigma'(r)$). The resulting equation for the pdf is:

$$p_{G|R}(g|R = r, Ab) = \frac{A'(r)}{\sqrt{2\pi}\sigma'(r)} e^{-\frac{(g-\mu'(r))^2}{2\sigma'(r)^2}}, p_R(r, Ab) = \varphi'(r)\,.$$

Table 5.1 summarizes the total number of parameters and their values for both modeled distributions: the no-abandon and the abandon.

5.4.2 Model Learning

Once the model was designed, its parameters were learned from two-thirds of the data. The two prior probabilities $P(Ab)$ and $P(\overline{Ab})$ were also learned by considering how much data were of both distributions as in the Chapter 4. The number of occurrences of the abandon distribution was divided by the total number of occurrences of both the abandon and the no-abandon distributions, given us $P(Ab) = 0.84$ and $P(\overline{Ab}) = 0.16$. Then, the total number of learned parameters was 14 (6+7+1). Figure 5.12 shows the two posterior probabilities $P(\overline{Ab}|g, r)$ and $P(Ab|g, r)$ after learning in order to represent the decision frontier between the two classes. As Fig. 5.12 shows, the intersection is oblique. There is dependency between *Rank* and *Gap*: at the beginning of processing

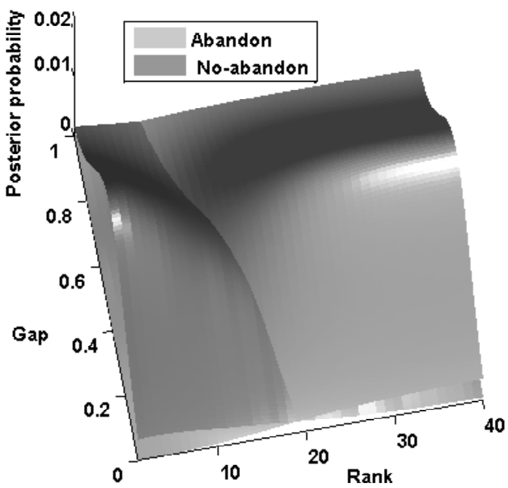


Figure 5.12: The posterior probabilities $P(\overline{Ab}|g, r)$ and $P(Ab|g, r)$ in the $Gap \times Rank$ space.

the paragraph (low values of the *Rank*), there should be a high difference between the two paragraphs to make the decision. However, after more fixations have been made, that difference could be lower to decide to abandon the paragraph. For instance, at rank 10, a *Gap* of 0.86 is necessary to stop reading, whereas at rank 15, a value of 0.42 is enough. The frontier is rather linear as is shown in Fig. 5.13 which is very convenient for our modeling purposes. In this way, we have a simple modeling function with a minimum number of parameters. Thus, the frontier can be approximated by the following equation in the $Gap \times Rank$ space that it is included in the computational model:

$$Gap_0 = -0.090 \times Rank + 1.768\,.$$

The model constantly computes the Gap_0 value while processing a paragraph, increasing the *Rank* value. As soon as the current *Gap* value is greater than Gap_0, the decision is to stop reading the paragraph. For example, Fig. 5.14 shows the cosine evolution of an experimental scanpath (dotted line) and the model frontier (straight line). The scanpath is composed of 20 fixations but the model decide to abandon the paragraph at fixation number 19 when the *Gap* value is greater than Gap_0 computed by the model. However, we can observe that participant made 1 more fixation than the model before to stop reading the paragraph.

Once the model parameters have been set, the evaluation of the model is discussed now.

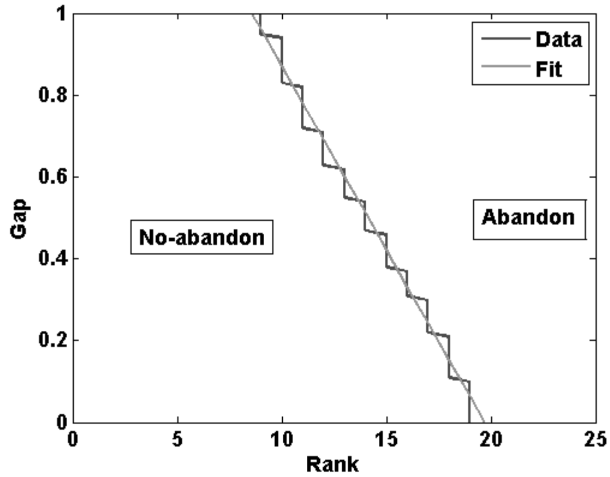


Figure 5.13: The Bayesian frontier between both the abandon and the no-abandon cases. Experimental data and its linear approximation are shown.

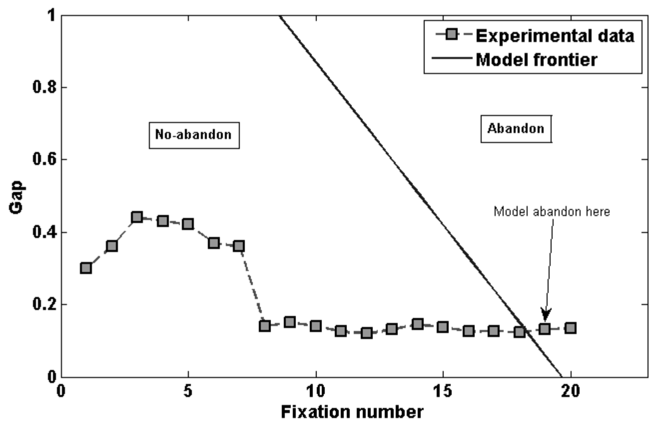


Figure 5.14: Cosine evolution of an experimental scanpath and the model frontier.

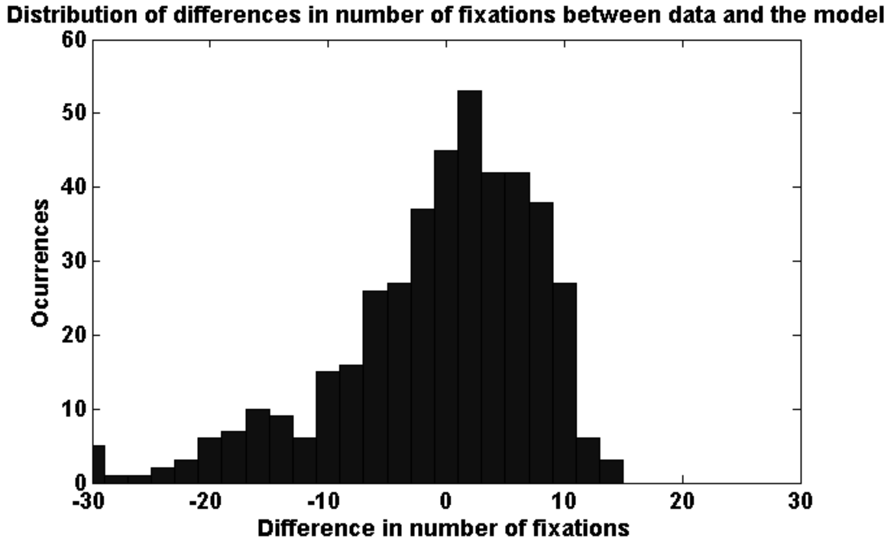


Figure 5.15: Distribution of the differences between the number of fixations (ranks) made by the participant and the model.

5.4.3 Model Performance

The model is tested it on the remaining one third of the data. The two cases considered were: 1) model stops reading either before or at the same time of participant and 2) model stops reading after the participant.

The average difference between the ranks at which model and participant stopped reading was computed only for the case number 1 where the model is more reactive than participant. We got a value of 5.17 (SE=0.22). It means that, in the average, the model stopped reading 5.17 fixations earlier than participants. This was observed in 59% of the cases and in 41% of the cases the model stopped reading after the participant. This situation can be observed on the distribution of the differences between the number of fixations made by the participant and the model as is showed in Fig. 5.15. The plot shows a slightly tendency of the model to stop reading a paragraph either before or at the same time of participant (left side of the plot). A total of 427 scanpaths were considered in the testing set.

For comparison purposes with the random model, the best p value found such as the random model had the smallest average difference with participant ranks was for $p = 0.2$. That smallest average difference was of 11.84 (SE=0.73). So, the results presented in the current section indicates that our model is about twice (in average) better than the best random model (5.17 fix. vs 11.84 fix.).

5.5 Conclusion

So far, we have presented a computational cognitive model that is able to replicate the average people's behavior, when they are faced to a task of quickly deciding if a paragraph presented is more related or not to a given goal than another paragraph processed previously. Two variables are involved: the rank of the fixation ($Rank$) and the difference of semantic similarity between the two paragraphs with the search goal (Gap). We proposed a simple linear threshold to account for that decision. The results are promising but we think that there should be a way to improve them. In the next chapter, some ideas that could help to improve the results are going to be presented.

Chapter 6

General Discussion

6.1 Discussion

We would like to start our discussion by comparing the different models presented so far in the previous chapters.

Fig. 6.1 shows the superposition of the different model frontiers in the $Cos \times Rank$ space learned from the Experiment 1. This experiment was concerned with Situation 1 (one goal, one paragraph) described in Chapter 3 (p. 39) and whose modeling work was presented in Chapter 4, where participants were asked to read a paragraph with the task of quickly deciding if it is related or not to a given goal. The three straight-lines correspond to the original model frontier without any participant's delay ("Original"), the model frontier by using the participant's delays learned from *ramp* words ("Delay-ramp") and the model frontier by using the participant's delays learned with the KL divergence ("Delay-KL"). As we can notice, the position of the model for the "Original" case is shifted to the right along the x-axis (*Rank*) compared to the others models when participant's delay are considered. This situation was expected because the idiosyncrasy of each participant is not considered in the "Original" case. All the extra fixations needed by participants to confirm their decision are taken into account for the modeling work, resulting in a curve shifted to the right because the *Rank* of abandoning a paragraph is higher.

The straight-lines corresponding to the "Delay-ramp" and "Original" models seem to be parallel. This is not the case for the 'Delay-KL" and "Original" models. They have a different slope. Here, we can notice that for low values of *Cos* (y-axis) and high values of *Rank* (x-axis), the "Delay-KL" model appears shifted to the left with respect to the "Original", but for high values of *Cos* and low values of *Rank*, the "Delay-KL" model appears shifted to the right. This situation could be due to the way that delays were computed by using the KL divergence. As we have explained in Chapter 4 (p. 84), the delays are computed by selecting the distribution of one participant as reference. All the participant's delays are dependent of this reference distribution. When we have a low KL divergence between the distribution of one participant and the reference distribution, the delay value learned for this particular participant is also low because we do not need to apply a big shift in order to minimize its divergence with respect to the reference distribution. This also applied in the reverse way when we have a high divergence between a participant's distribution and the reference distribution. From Fig. 6.1, we

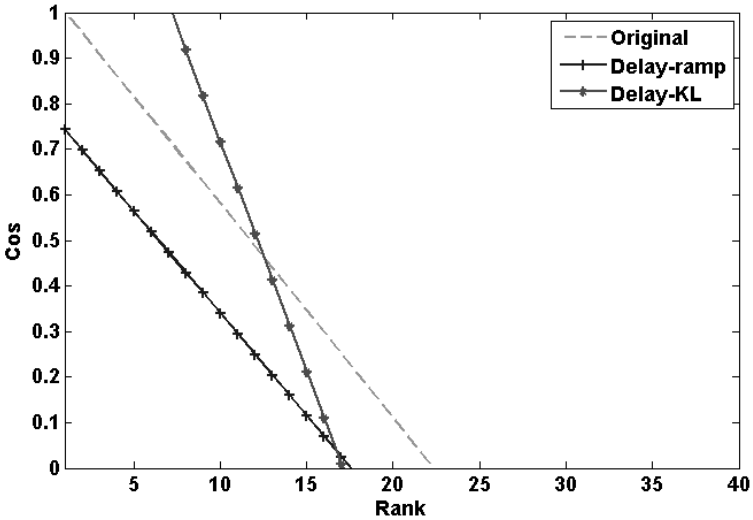


Figure 6.1: Different models learned from the Experiment 1 in the $Cos \times Rank$ space.

can say that the “Delay-KL” model is shifted to the left of the “Original” for low values of Cos and high values of $Rank$, because we suspected that, there, participant’s delays applied were higher and they were lower for high values of Cos and low values of $Rank$. However, a more detailed investigation could be done. One way to do that could be to analyse all the participant’s distribution in the $Cos \times Rank$ space and compare them to the corresponding participant’s delay. However, we were not interested about that in the current thesis because we considered that other important aspects about our work should be also discussed here.

Table 6.1 shows the average difference between the ranks at which model and participant stopped reading a paragraph for the three versions of the model. The total number of parameters used to fit both posterior probabilities corresponding to the no-abandon ($\overline{Ab}$) and the abandon (Ab) distributions and the number of parameters used to fit the model frontier are also shown. The “Original” model is the worst one, with an error of 4.7 fixations compared to the human data. There is not a significant difference of performance between both the “Delay-ramp” and “Delay-KL” models (t-test, $p>0.3$), but as we have argued before, the case of the model learned by considering the *ramp* word should not be considered as a model improvement. In fact, we have mainly two arguments about that. Firstly, we remind that a *ramp* is characterized by two indexes, the beginning of the *ramp* and the end of the *ramp*. In our approach, we used the end of the *ramp* but we could had used the beginning of the *ramp* or any index in between. There was no any particular reason to select this as the *ramp* reference. As we can remember, the idea was that we suspected that fixations on special words would trigger

Model	Diff. in ranks	Posterior prob.	Frontier
Original	4.7 (SE=0.72)	$6(\overline{Ab}) + 8(Ab) + 1(\text{prior}) = 15$	2 (linear)
Delay-ramp	3.0 (SE=0.5)	$8(\overline{Ab}) + 10(Ab) + 1(\text{prior}) = 19$+delays	2 (linear)
Delay-KL	3.05 (SE=0.83)	$8(\overline{Ab}) + 8(Ab) + 1(\text{prior}) = 17$+delays	2 (linear)

Table 6.1: Average difference between the ranks at which model and participant stopped reading a paragraph for the three versions of the model of the Experiment 1. Number of parameters used are also indicated.

Model	Diff. in ranks	Posterior prob.	Frontier
Original	5.17 (SE=0.22)	$6(\overline{Ab}) + 7(Ab) + 1(\text{prior}) = 14$	2 (linear)

Table 6.2: Average difference between the ranks at which model and participant stopped reading a paragraph for the Experiment 2. Number of parameters used are also indicated.

the decision to stop reading a paragraph and we proposed three candidates to use as special words: *step*, *target* and *ramp* word. In the case of the *step* and *target* words, both are well-defined by only one word each but unfortunately in our modeling work, any of them allowed us to have a good result. Secondly, the total number of parameters needed by the "Delay-ramp" model (19+1=20 parameters) is higher than the corresponding to the "Delay-KL" model (17+1=18 parameters). Then, the results got from the "Delay-KL" model was the only improvement achieved. In average, the "Original" model decides to stop reading a paragraph 4.7 fixations earlier than participants and the "Delay-KL" model does it in only 3.05 fixations earlier than participants. However, the "Delay-KL" model (19+1=20 parameters) uses more parameters than the "Original" model (17 parameters).

Fig. 6.2 shows the model frontier learned from the Experiment 2 concerned with SITUATION 2 (one goal, two paragraphs) described in Chapter 3 (p. 39) and whose modeling work was presented in Chapter 5, where participants were asked to read a paragraph with the task of quickly deciding whether it is better related to a given goal than another paragraph processed previously. Table 6.2 shows the average difference between the ranks at which model and participant stopped reading the paragraph. The number of parameters used as in the previous case is also indicated.

Making a comparison between the models of the SITUATION 1 and SITUATION 2 is not straightforward, because the first one occurs in the *Cos* × *Rank* space and the second in the *Gap* × *Rank* space. Even if both variables *Cos* and *Gap* range from 0 to 1, *Cos* implies a semantic similarity between the words actually processed in one paragraph and the search goal and *Gap* implies a difference of semantic similarities between words actually processed in two paragraphs with the search goal. Moreover,

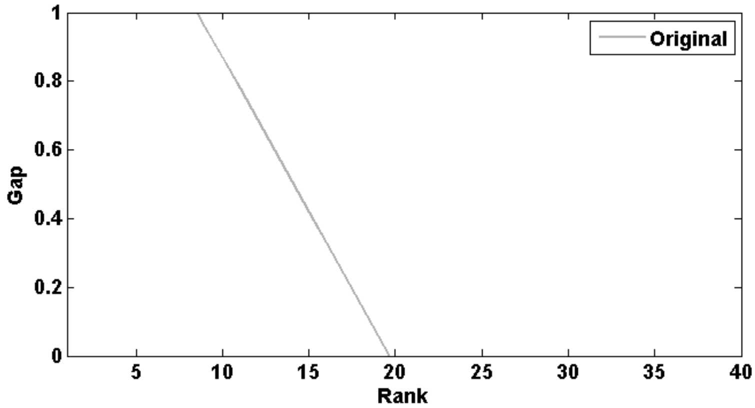
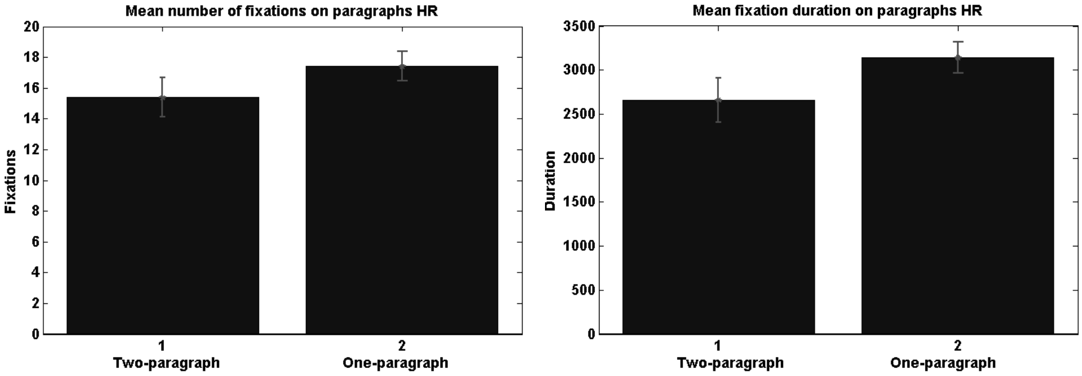


Figure 6.2: Model learned for the Experiment 2 in the *Gap* × *Rank* space.

during the Experiment 1, people are encouraged only to make a decision about the relatedness of a paragraph with the search goal whereas in the Experiment 2 people are encouraged not only to make a decision about the relatedness of a paragraph with the search goal but also a comparison between two paragraphs in order to choose the best one. However, we made the comparison of two basic statistics between the two experiments: the mean number of fixation and its duration while processing paragraphs of type HR. Results are shown in Fig. 6.3a and 6.3b respectively. For the Experiment 1, we got that people made in average about 17.40 fixations (SE=0.94) in each paragraph and they spent about 3139.8 ms (SE=177.9). For the Experiment 2, people made in average 15.38 fixations (SE=1.26) and they spent about 2656.7 ms (SE=251.9). The very light difference maybe due to the fact that, during the Experiment 2, people know that they need to make a comparison between two paragraphs and this situation might encourage them to try to do it faster than in the case of the Experiment 1 where only one paragraph is presented at time. However, there is not a significant difference neither on the average number of fixations (t-test, $p>0.25$) nor the mean duration (t-test, $p>0.15$). Thus, we could say that in average people made the same number of fixations on each paragraph and they spend the same time visiting the paragraphs across both experiments.

Now, we are going to discuss about the complexity of the models presented in the current thesis. The aim of the thesis was to build computational cognitive models able to replicate in average people's behavior while they are looking for information and not to reproduce their behavior perfectly. For that reason, during the conception of our models the idea was to have a low number of parameters in order to avoid the overfitting and the noisy data. Overfitting generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. A model which has been overfitted will generally have poor predictive performance, as it can exaggerate



(a) Number of fixations (b) Fixations duration

Figure 6.3: a)Mean number of fixations and its overall duration while processing a paragraph HR for both experiments.

minor fluctuations in the data. For example, if the number of parameters is the same as or greater than the number of observations, a simple model or learning process can perfectly predict the training data simply by memorizing the training data in its entirety, but such a model will typically fail drastically when making predictions about new or unseen data, since the simple model has not learned to generalize at all. Our models were built as simple as possible. Even there was not a rigourous work at minimizing the number of parameters or at reducing the complexity of the fitting functions used along the modeling work, we tried to use the minimum number of parameters with simple equations at two levels. In a first level, this consideration was done at the parametrisation stage for both the no-abandon and the abandon distributions as it was showed in Chapter 4 and 5. Some parameters of both distributions were fitted with simple linear, exponential or logistic equations but all the time, we tried to choose the simplest as possible. In a second level and once that both distributions were modeled, the model frontier was fitted all the time with simple linear equations. However, a more rigourous work should be done in such direction if we want to evaluate the impact over the model performance of using different complexity equations with different number of parameters for fitting the data.

Some of the benefits, limits and drawbacks of the computational cognitive models are given now.

6.2 Benefits, limits and drawbacks of our modeling work.

We start by focusing on the benefits of our modeling work. Two different situations of information search have been studied along the current thesis and one model is proposed for each one. With the two models, we are able to make predictions about the moment at which participants would decide to stop reading a paragraph when they are faced to information search tasks like those that were presented here.

However, and as it was expected, our models do not replicate exactly people's behavior as ideal models would do. Some of their limitations are discussed now. We should remember that the evaluation of our model's performance was done by following experimental scanpaths of testing sets. It was done in that way because our models only decide the *time* at which participants will stop reading a paragraph but they do not have a mechanism to decide *where* participants will make new fixations. Then, the first limitation and the most important is that given a new paragraph, our model will suppose that the paragraph is processed in a linear way and this is not exactly the case in an information search task. We did not make a lot of effort in this direction because we considered that it was more important to have a functional model able to predict the *time* at which participants will stop reading a paragraph. However, different non-linear strategies of exploring texts, such as skimming as is shown in Carver (1990), have been modeled in our team (Lemaire et al., 2011) and could be connected to the model in the future.

Our models are completely based on LSA (Landauer et al., 2007) as a model of semantic memory. Human judgments of semantic associations are mimicked by LSA. As an automatic tool, LSA has inherent drawbacks due to the method of computing similarities as it was explained in Chapter 3 (p. 41). First, although the 300 dimensional space is large enough to represent different meanings for polysemous words, it could be the case that different meanings are mixed if the context is not rich enough to disambiguate them. A goal composed of 2 to 3 words is not a large enough context to ensure that the correct meaning is always selected. A second one could be the fact that LSA assumes that words and documents form a joint Gaussian model while a Poisson distribution has been observed (Hofmann, 1999). We also want to mention that LSA is not capable to solve the well-known problem of BOW (bag of words) because a sentence is represented as an unordered collection of words, disregarding grammar and even word order. We can imagine that we extract one sentence of one of our paragraphs used in our experiments and we compute its cosine similarity with the corresponding goal. Then, we could reorder all the words and remove some words (prepositions, articles) of this first sentence without preserving any grammar rule. Surely, the new sentence will not have the same meaning but the computation of its cosine similarity with the goal will be the same as before. In this case, the LSA method treats both sentences as the same. This could be also the case when a participant make only a few fixations on a paragraph by picking here and there some words and trying to get a perception about the relatedness of the paragraph with the goal.

In spite of the drawbacks of LSA method, we got good model's performances but a further investigation should be done in order to know in more details the impact of these limits in our modeling work.

Another limitation is that in general during the conception of the models, our analysis was only concerned with paragraphs of type HR and not paragraphs of type MR or UR. This consideration was made only for simplifying purposes because we showed that deciding that a paragraph is related to the goal (HR) is different from deciding that it has nothing to do with the goal (UR). In the case of paragraphs MR, things are still more complicated due to their neutral relatedness to the search goal. Moreover, during the construction of the model for the Experiment 1, a clustering process was also applied in order to select the material according to its cosine evolution and at the end only the paragraphs clustered as *ramp* paragraphs were considered for our modeling work. Future developments of the current work could be to extend the range of paragraphs that our models are able to tackle.

Finally, we would like to present some real applications of the proposed models.

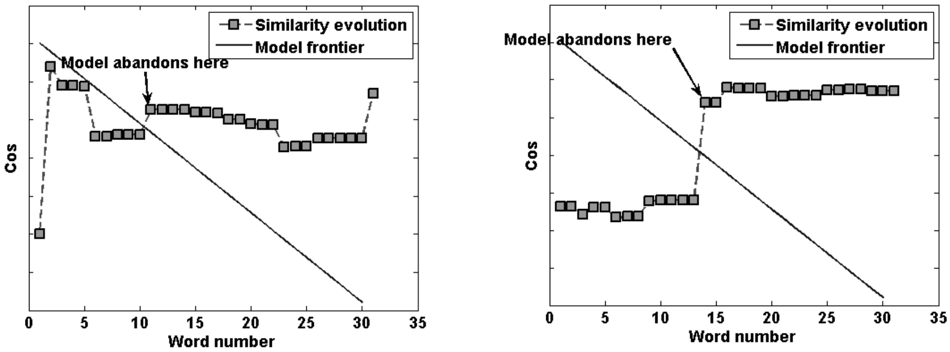
6.3 Practical applications.

Knowing where a user is looking at while searching for information is crucial for evaluating the usability of Web sites and contribute to the design of Web pages (Baccino et al., 2005). For that, it is important to have powerful simulation tools (models) to explain and to simulate how the visual exploration is made on Web pages containing both texts and images and as a consequence to provide fruitful information to any interface designers (Chanceaux et al., 2009). Web designers are asked to build pages as attractive as possible in order to have the visitor's attention. This attractiveness could be at the level of the layout interface, the visual information (images) or texts presented among others. Our model can be used to perform an automatic selection of several textual versions of a content to be presented into a homepage (as in a simplified news web portal). In this case, such an automatic tool is used as a virtual participant. Our model could give an answer to the following question: given a text and its corresponding headline (it can see it as a search goal), how the information should be arranged into the paragraph? Will the effect to participant's attention be the same by presenting the information related to the headline at the very beginning of the paragraph rather that at the end or at the medium of the paragraph? Suppose, that we have two different versions of text concerning the theme "Aide aux réfugiés" as below:

Text 1: "Le HCR prévoit des campagnes d'information de masse dans les camps pour sensibiliser au danger des mines. L'état de l'unique route ne peut supporter un afflux simultané de tous les réfugiés."

Text 2: "L'état de l'unique route ne peut supporter un afflux simultané de tous les réfugiés. Le HCR prévoit des campagnes d'information de masse dans les camps pour sensibiliser au danger des mines."

The texts are composed of two sentences containing the same words. Sentences have



(a) TEXT 1 (b) TEXT 2

Figure 6.4: Goal-similarity evolution of the TEXT 1 and TEXT 2 to the the theme “Aide aux réfugiés”.

been only swapped. Then, the overall goal-similarity of both texts to the theme “Aide aux réfugiés” is the same (0.74 in this case). However, the goal-similarity evolution should be different and can be computed. We assume a linear way of text reading. We can use our model presented in Chapter 4 to decide the time at which a user will stop reading each text.

Fig. 6.4a and 6.4b show the goal-similarity evolution of TEXT 1 and TEXT 2 to the theme “Aide aux réfugiés”. The straight-line corresponds to the model frontier (Eq. 4.4.2) presented in Chapter 4 (p. 70). In the case of the Fig. 6.4b, our model decides to stop reading at fixation number 14 whereas in the case of the Fig. 6.4a, our model needs to do only 11 fixations. These results suggest that we could consider that TEXT 1 to be better than TEXT 2 for the current goal because when it is presented to a participant, the virtual participant only needs to do in average 11 fixations to have a perception about its relatedeness to the theme “Aide aux réfugiés”.

We can imagine an application for our 2-paragraph model presented in Chapter 5. This model runs in the *Gap* × *Rank* space. *Gap* represents the difference of semantic similarities between each of two paragraphs and a search goal. In this way, the model could be used to predict not only what users would do when reading a single paragraph, but what users would do with a page of several paragraphs. The model “reads” a paragraph and abandoned it somewhere. Then, it could “read” another one (the current) having the previous one in mind. The current paragraph is likely to be abandoned when the *Gap* value reaches the threshold value. Here, the *Gap* is computed as the difference of semantic similarities between the previous paragraph and the current paragraph to the search goal. That way, a web designer could study how users would behave on his web page, given a goal. For example, the web designer of a car insurance company home page could test whether it takes a long time or not to users to find a specific information, say about baby car seats.

Appendix A
Appendixes

A.1 Experiment 1

A set of 30 themes were created for this experiment. 6 paragraphs (hr_1, hr_2, mr_1, mr_2, ur_1 and ur_2) were considered for the stimuli pages. The description of all the paragraphs is presented later.

A.2 Experiment 2

A subset of 20 themes were considered for this experiment. 7 paragraphs were managed for the stimuli pages (hr_1, hr_2, mr_1, mr_2, ur_1, ur_2 and ur_3). The description of all the paragraphs is also presented below.

A.3 Text of the stimuli pages

Texts of the paragraphs used. For each theme, the 7 paragraphs are listed: the 2 paragraphs with strong semantic association to the theme (HR: hr_1 and hr_2), the 2 paragraphs with medium association (MR: mr_1 and mr_2) and the 3 paragraphs without any association to the theme (UR: ur_1, ur_2 and ur_3). The cosine value of the similarity is also indicated (Cos).

Theme	Type	Cos	Text
Aide aux réfugiés	hr_1	0.805	"Lorsque la vague des réfugiés déferlait sur le camp improvisé, les microbus et les tracteurs mobilisés par l'organisation macédonienne El Hilal assuraient un mini-pont humanitaire qui permettait d'éviter le pire."
Aide aux réfugiés	hr_2	0.739	"L'état de l'unique route ne peut supporter un afflux simultané de tous les réfugiés. Le HCR prévoit des campagnes d'information de masse dans les camps pour sensibiliser au danger des mines."

Aide aux réfugiés	mr_1	0.166	"Le catalogue des principes du retour au calme à la frontière, établi par les ministres, reprend l'essentiel des cinq conditions posées par les gouvernements occidentaux pour un arrêt des frappes."
Aide aux réfugiés	mr_2	0.176	"Les habitants de Tirana compatissent aux malheurs de leurs frères kosovars. Ils se déclarent ulcérés par l'enchaînement d'événements qu'ils ressentent comme une profonde injustice et reprochent à l'UCK d'avoir lancé des opérations armées."
Aide aux réfugiés	ur_1	-0.011	"Entre le 31 décembre à minuit et jusqu'à la fin du mois de janvier, les clubs de football sont autorisés à renouveler leur effectif."
Aide aux réfugiés	ur_2	-0.009	"Le chef du gouvernement a estimé que la croissance économique de la France était sur la bonne voie contrairement aux prévisions des experts. La croissance du produit intérieur brut est étonnante."
Aide aux réfugiés	ur_3	-0.036	"Les enfants jouent au foot avec des balles de tissus noués, nagent dans la rivière, s'entraînent au tir à l'arc, loin du regard préoccupé d'un adulte. Ils vont aussi à l'école."
Allocation familiale	hr_1	.346	"Les familles ayant des jumeaux bénéficient des mêmes prestations familiales que les autres, mais les congés maternité sont plus longs de 18 semaines."
Allocation familiale	hr_2	.320	"L'allocation de rentrée scolaire concerne 350 000 familles modestes ayant un enfant scolarisé. Cette prestation pourrait être prochainement étendue."
Allocation familiale	mr_1	.189	"La moitié des tout-petits sont gardés au foyer par un parent, la maman dans la quasi-totalité des cas. Seuls 8
Allocation familiale	mr_2	.140	"Mme aubry a annoncé que l'objectif national des dépenses d'assurance maladie voté par le parlement sera dépassé d'environ 2 milliards d'euros."
Allocation familiale	ur_1	.011	"Le bombardier Spirit que les visiteurs du salon du Bourget ont entraperçu en juin, le temps d'un très bref passage au-dessus de leurs têtes, est une drôle de machine volante pour les non initiés."
Allocation familiale	ur_2	-0.035	"La pianiste américaine Rosalyn Tureck fait son entrée, à quatre-vingt-quatre ans, dans le répertoire discographique de Deutsche Grammophon. Son interprétation des variations Goldberg de Bach va la faire connaître au grand public."

Allocation familiale	ur_3	-.076	"Banlieue, violence et parfois urgence... Ces mots constamment accolés réapparaissent en force à chaque fois qu'il est question dans l'actualité de chauffeurs de bus agressés, de véhicules incendiés.
Art contemporain	hr_1	.507	"On se souvient du fameux symbole du dollar revu par Andy Warhol. Le pop-artiste américain avait d'un signe simple amalgamé l'art et l'argent pour la plus grande joie de ses collectionneurs."
Art contemporain	hr_2	.576	"Le musée d'art moderne de la ville de Paris présente jusqu'au 18 avril soixante-neuf peintures de Mark Rothko, un des grands classiques de la peinture abstraite de l'après-guerre aux Etats-unis."
Art contemporain	mr_1	.230	"En raison des travaux entrepris dans l'aile nord du château de Versailles, les salles des croisades et le circuit du XVIIe siècle, qui font partie du musée de l'histoire de France, seront fermées."
Art contemporain	mr_2	.212	"Les galeries d'anatomie comparée et de paléontologie du musée national d'histoire naturelle sont de nouveau ouvertes au public depuis le 18 décembre, après trois mois de travaux."
Art contemporain	ur_1	.005	"Un rapport de l'inspection générale des affaires sociales a mis le feu aux poudres en dénonçant une série de dysfonctionnements dans la gestion du comité national contre le tabagisme."
Art contemporain	ur_2	-.021	"Une information judiciaire pour blessures involontaires commises par négligences a été ouverte après la chute d'un enfant de six ans, tombé d'un wagon non équipé de fermetures de sécurité du train Toulouse-Paris."
Art contemporain	ur_3	-.015	"Un jeune homme de dix-neuf ans a été mis en examen pour violence volontaire avec arme par destination pour avoir lancé son chien sur trois ouvriers d'un chantier de Colombes."
Associations humanitaires	hr_1	0.598	"La détresse et la souffrance des réfugiés largement médiatisées, suscitent d'immenses élans de générosité. Des convois d'aide humanitaire constitués par des bénévoles de différentes organisations convergent vers les camps de réfugiés."
Associations humanitaires	hr_2	0.521	"Les collectivités locales sont encouragées à coordonner leur action en faveur des réfugiés. Elles envisagent de monter une plate-forme pour assurer la traçabilité de l'acheminement des colis humanitaires."

Associations humanitaires	mr_1	0.185	"Il s'agit de nos citoyens et le devoir de n'importe quel gouvernement est de défendre les intérêts de ses citoyens. Des mesures sont prises mais également une coopération avec les organisations internationales."
Associations humanitaires	mr_2	0.206	"Le problème que devra gérer le gouvernement est de canaliser le retour des sinistrés dans les familles d'accueil. Le préalable est leur identification car ils n'ont plus de pièce d'identité."
Associations humanitaires	ur_1	-0.028	"La déflagration s'est produite au passage d'un véhicule militaire sur une route du centre-ville, à une centaine de mètres d'une base de l'armée turque, selon la police."
Associations humanitaires	ur_2	-0.001	"La nageuse a décroché sa deuxième médaille d'or aux championnats d'Europe de natation en petit bassin en s'imposant dans son épreuve de prédilection. La Française l'a emporté facilement, reléguant sa grande rivale."
Associations humanitaires	ur_3	0.087	"Les élections de l'Iowa ont été sans doute les plus disputées de l'histoire de cet État rural. La participation à ces assemblées a dépassé tous les espoirs prévus par les politiques."
Chasse aux oiseaux	hr_1	.425	"Dans le Pas-de-Calais, les chasseurs réclament le droit de continuer à chasser dans les huttes. La France risque pourtant une lourde amende si elle contrevient aux directives européennes qui protègent les oiseaux migrateurs."
Chasse aux oiseaux	hr_2	.345	"Plusieurs tribunaux administratifs ont sommé les préfets de treize départements de prendre des arrêtés de fermeture de la chasse aux oiseaux migrateurs au 31 janvier, réduisant de trois semaines la durée de la chasse."
Chasse aux oiseaux	mr_1	.104	"L'ortolan, petit oiseau très prisé par les gourmets du Sud-ouest, appartiendra bientôt aux espèces protégées. Un décret a été signé qui représente le premier acte d'un compromis global entre écologistes et chasseurs."
Chasse aux oiseaux	mr_2	.106	"Les petites boules de graisse enrobées de graines diverses sont très appréciées des mésanges qui les prennent d'assaut aux premières rigueurs de l'hiver."
Chasse aux oiseaux	ur_1	.023	"Les agences spatiales françaises et japonaises ont décidé de créer des groupes de travail sur le suivi des risques naturels. L'un d'eux sera chargé de réfléchir à une meilleure utilisation des données spatiales."

Chasse aux oiseaux	ur_2	.037	"L'édition d'art offre aux jeunes éditeurs ambitieux de grandes satisfactions : on peut y avoir le sentiment de faire les livres de A à Z, comme un producteur de Hollywood ses films."
Chasse aux oiseaux	ur_3	.016	"La place financière parisienne est privée de deux piliers majeurs : des fonds de pension qui draineraient une épargne investie en actions et des banques importantes ayant un pied aux Etats-unis."
Chef de la Russie	hr_1	0.730	"A Moscou, les communiqués triomphants se succèdent. Dans le Daghestan, les rebelles sont complètement encerclés, leur chef militaire est blessé, les islamistes ont subi de lourdes pertes."
Chef de la Russie	hr_2	0.464	"Depuis le limogeage surprise du précédent chef du gouvernement russe après cinq ans de stabilité, le président tente sans succès de maîtriser sa succession. Les prétendants sont nombreux."
Chef de la Russie	mr_1	0.138	"Ces combats, qui ne remettent pas en cause la suprématie des talibans sur l'Afghanistan, prouvent que l'état-major veut et peut maintenir la pression sur ses adversaires."
Chef de la Russie	mr_2	0.172	"Une bombe de forte puissance a explosé dans le centre de la capitale sur le passage du président tchétchène. L'attentat n'a pas encore été revendiqué, indique Moscou."
Chef de la Russie	ur_1	-0.011	"Il n'est point de bon mercato sans un feuilleton pour l'animer. Les transferts des deux footballeurs n'ont pas électrisé les foules en dehors du premier cercle des admirateurs de l'un et de l'autre."
Chef de la Russie	ur_2	0.005	"Le tribunal administratif de Paris a annulé la décision du Conseil de Paris d'étendre à des communes de banlieue le système de vélos en libre-service."
Chef de la Russie	ur_3	-0.045	"Son arc en bandoulière, le chasseur avance dans la forêt. L'Indien traque un singe ou des oiseaux, proies faciles pour ses flèches couronnées de plumes rouges et enduites de curare."
Conflit en Irak	hr_1	0.630	"Les États-Unis ont profité de la commission spéciale de l'ONU chargée de désarmer l'Irak pour espionner Bagdad. Les responsables américains admettent avoir profité des informations militaires pouvant aider à renverser le régime."
Conflit en Irak	hr_2	0.755	"L'Irak a demandé à l'ONU d'empêcher les survols américains et britanniques et de permettre à ses avions civils de transporter des scientifiques étrangers. L'Irak est soumis à un embargo aérien depuis 1990."

Conflit en Irak	mr_1	0.227	"Comme en écho à cette assurance retrouvée depuis la fin de la guerre, la commission des affaires étrangères du Sénat a fixé une date pour le début des auditions et la nomination qui était bloquée depuis dix mois."
Conflit en Irak	mr_2	0.181	"L'arsenal militaire du pays s'est enrichi d'armes à micro-ondes susceptibles de détruire les circuits informatiques et les équipements de télécommunications. Les militaires veulent les garder pour s'en servir contre les infrastructures ennemies."
Conflit en Irak	ur_1	0.028	"Depuis l'interdiction de fumer, c'est l'affluence d'un dimanche. D'un côté, on fait la queue pour acheter des cigarettes, de l'autre on patiente avec le bordereau des courses hippiques du jour."
Conflit en Irak	ur_2	-0.026	"L'entraîneur dirige l'équipe féminine seulement cinq ans après avoir été champion du monde comme coach adjoint des hommes. Il a ses petites habitudes. Le rituel commence pendant l'hymne russe."
Conflit en Irak	ur_3	-0.053	"Les esquimaux vivent dans les régions polaires très froides. En été, ils pêchent et ils chassent. Ils habitent dans des tentes faites de peaux de phoque. En hiver, les esquimaux logent dans des maisons de bois."
Conflit israélo-palestinien	hr_1	0.044	"L'artillerie israélienne a massivement bombardé des régions supposées être des bastions du Hezbollah au Liban. Ces tirs sont une riposte à une série d'attaques meurtrières de la milice intégriste Chiite contre Israël."
Conflit israélo-palestinien	hr_2	0.157	"Proche-orient : le chef du gouvernement a promis de retirer l'armée israélienne du Liban sud dans un délai d'un an. Il envisage de négocier simultanément avec la Syrie qui entretient un dispositif militaire propice au conflit avec Israël."
Conflit israélo-palestinien	mr_1	0.058	"Le chef du gouvernement est accusé des relations tendues avec les États-Unis, de l'isolement d'Israël dans le monde, du chômage en montée vertigineuse, et de la détérioration de toutes les normes de gouvernement démocratique."
Conflit israélo-palestinien	mr_2	-0.009	"Pour l'instant, les télécommunications palestiniennes sous-traitent l'accès à l'international à une compagnie privée israélienne. Les réseaux sont au point et on trouve maintenant des cabines téléphoniques publiques à carte."

Conflit israélo-palestinien	ur_1	-0.004	"Le juge administratif estime que cette extension du système de locations de vélos en libre-service constitue un nouveau marché. Il fournit pour les communes un réel service de bicyclettes en libre-service."
Conflit israélo-palestinien	ur_2	-0.030	"Les laboratoires ont ainsi réalisé qu'au premier coup dur, leur existence même était menacée. Les médicaments nouveaux n'étaient plus assez nombreux pour compenser la chute de leurs brevets dans le domaine public."
Conflit israélo-palestinien	ur_3	0.030	"Ribéry respire la joie de vivre, il ne se pose aucune question sur sa manière de jouer, il ne peut faire que du bien. Franck deviendra quelqu'un d'important."
Course des voiliers	hr_1	0.688	"Le monocoque Gartmore du Britannique a démâté dans la troisième étape de l'Around Alone, tour du monde à la voile en solitaire avec escales, entre Auckland et Punta del Este."
Course des voiliers	hr_2	0.448	"Le navigateur solitaire voulait gagner pour l'honneur. Un choix météo décevant et une grosse avarie dans le Pacifique ont eu raison de ses chances. Il a vécu un drôle de tour du monde."
Course des voiliers	mr_1	0.197	"La formule 1 a réclamé de ses pilotes qu'ils soient les plus rapides à tout moment, la discipline reine du sport automobile exige parfois de ses conducteurs qu'ils rongent leur frein."
Course des voiliers	mr_2	0.192	"L'équipe cycliste a retrouvé quelques couleurs sportives, alors que les noms de ses représentants figuraient plus souvent à la rubrique des affaires, au gré des avancées des enquêtes liées au dopage."
Course des voiliers	ur_1	0.025	"La Chine souhaite généraliser l'usage de l'injection mortelle pour les peines capitales au détriment de l'exécution par balle. C'est considéré comme plus humain dans toutes les juridictions intermédiaires."
Course des voiliers	ur_2	-0.021	"Le fondateur de la guérilla, en rébellion contre les autorités, ne fait aucune allusion à l'opération humanitaire visant à remettre trois otages qui a capoté en début de semaine dernière."
Course des voiliers	ur_3	0.066	"Le lancement d'Atlantis a été encore reporté en raison de problèmes récurrents de lisibilité du capteur. Celui-ci permet de connaître le taux de remplissage du réservoir d'hydrogène de la navette."
Croissance de l'économie	hr_1	0.919	"Le pacte de stabilité et de croissance pose des limites à l'utilisation du budget pour relancer une économie. Les économistes s'interrogent sur la pérennité de cette croissance qui est étonnante."

Croissance de l'économie	hr_2	0.855	"Une économie New Age est une économie dans laquelle la croissance de l'efficacité avec laquelle nous utilisons les ressources productives se trouve durablement accélérée, l'inflation maîtrisée et le cycle aboli."
Croissance de l'économie	mr_1	0.371	"La globalisation de l'économie est loin d'être achevée et, dans un contexte de fort chômage, la probabilité que s'imposent à nouveau les vieux mécanismes de hausse des prix est très faible."
Croissance de l'économie	mr_2	0.514	"Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens entraînant une accélération de la croissance, la stabilité des prix et un excédent de la balance des paiements courants."
Croissance de l'économie	ur_1	-0.012	"Dans les clairières, les Indiennes récoltent du manioc, du maïs, des bananes, des coeurs de palmier. Leurs paniers sont portés sur le dos et équilibrés par une lanière passée sur le front."
Croissance de l'économie	ur_2	-0.008	"Quatre jours de grains sont venus clore sa traversée du Pacifique. L'eau a gagné la bagarre et tout était trempé à bord. Le navigateur attend le moment idéal pour aller se reposer."
Croissance de l'économie	ur_3	0.011	"Voler a toujours été le grand délire de l'homme. Mais malgré ce désir et l'audace que requiert cette spécialité, il ne faut pas être totalement givré."
Déchets nucléaires	hr_1	.720	"La CRIIRAD a dénoncé la sous-évaluation par la Cogema de la radioactivité produite par des eaux de ruissellement d'anciennes mines d'uranium en Loire-atlantique."
Déchets nucléaires	hr_2	.358	"Plusieurs points de contamination radioactive ont été détectés dans un wagon transportant du combustible nucléaire usé provenant de la centrale de Bugey a indiqué EDF."
Déchets nucléaires	mr_1	.206	"La filiale déchets de Suez Lyonnaise des eaux a mis en garde mercredi 27 janvier les marchés d'une possible baisse de ses résultats à la suite d'une modification de la réglementation sur les décharges."
Déchets nucléaires	mr_2	.106	"Plus de 200 infractions à l'environnement ont été constatées dans le marais de Fontenay-le-vicomte lors d'une opération lancée par le parquet d'Evry : plusieurs constructions ont été recensées dans cette zone non constructible."
Déchets nucléaires	ur_1	-.014	"C'est la fête du vin à Luri en Haute-corse, un jour d'été, entre vendeurs de beignets au bruccio, de gâteaux à la farine de châtaigne et de charcuterie locale."

Déchets nucléaires	ur_2	.024	"Ils sont encore six candidats à vouloir tenter le tour du monde en ballon sans escale. Cette circumnavigation aérienne réclame autant de courage et d'audace que de moyens et de connaissances techniques."
Déchets nucléaires	ur_3	.023	"Le serial killer n'est pas un phénomène récent, mais ce n'est qu'au début des années 80 que la fiction policière va largement s'inspirer de ce thème, en littérature et au cinéma."
Décollage de fusée	hr_1	.342	"Le premier tir de l'année de la fusée Ariane-4 depuis le centre spatial guyanais de Kourou a été un succès : le lanceur européen a mis sur orbite un satellite de télécommunications."
Décollage de fusée	hr_2	.300	"La sonde américaine Stardust qui doit ramener sur terre des échantillons de la comète Wild-2 a été lancée de Cap Canaveral en Floride par une fusée Delta-2."
Décollage de fusée	mr_1	.132	"Un petit bimoteur qui effectuait la liaison Boston-Chicago a atteri en urgence à Pittsburgh hier en raison d'un défaut sur un des moteurs."
Décollage de fusée	mr_2	.168	"L'espace dans lequel évoluent les satellites autour de notre planète est jonché de milliers de débris de toutes sortes qui inquiètent les professionnels."
Décollage de fusée	ur_1	-.008	"Aujourd'hui, les grands noms de la pâtisserie française sont présents au Japon associés à des nippons. Il y a en outre pléthore d'excellents petits pâtissiers japonais de quartiers."
Décollage de fusée	ur_2	-.037	"Zinedine Zidane, grand artisan de la victoire de la France, en finale de la coupe du monde contre le Brésil, a été élu champion des champions du monde."
Décollage de fusée	ur_3	.021	"La fédération des consommateurs américains a demandé vendredi à Microsoft de rembourser 10 milliards de dollars aux consommateurs pour avoir pratiqué des prix trop élevés au cours des trois dernières années."
Délinquance juvénile	hr_1	.386	"Une quarantaine de véhicules ont été incendiés par des jeunes mineurs dans les banlieues de Strasbourg pendant la nuit de la Saint-Sylvestre. Des affrontements avec les forces de police ont également eu lieu."
Délinquance juvénile	hr_2	.318	"Une soixantaine de jeunes de la Courneuve ont incendié des voitures, brisé des vitrines et s'en sont pris aux policiers auxquels ils attribuent la responsabilité de la mort d'un des leurs."

Délinquance juvénile	mr_1	.141	"A Odense, au Danemark, chaque policier doit rester au moins un an dans un quartier populaire de la ville. Il passe un tiers de son temps à patrouiller à pied."
Délinquance juvénile	mr_2	.147	"La société HLM 3F a organisé une formation de trois jours réservée aux gardiens d'immeubles. L'objectif est de permettre aux personnels de terrain de garder la distance nécessaire dans les situations de violence."
Délinquance juvénile	ur_1	.044	"Après l'Amérique du Sud où il continue de nourrir les populations, le maïs a conquis le reste du monde où il est surtout utilisé pour engraisser les troupeaux."
Délinquance juvénile	ur_2	.030	"Les grands constructeurs automobiles cherchent à rajeunir leurs effectifs. La moyenne d'âge des salariés de l'automobile est de trente ans au Japon alors qu'elle atteint quarante-cinq ans chez Renault."
Délinquance juvénile	ur_3	.008	"Chez Vivendi, la direction générale n'impose aucune norme salariale à ses filiales. Par exemple, dans le pôle Eau, les salaires sont indexés sur la valeur du point de la fonction publique."
Faiblesse du dollar	hr_1	0.712	"L'Euro poursuivait sa remontée face au dollar, vendredi dans les premiers échanges entre bourses. Soutenu par les bonnes perspectives de croissance économique et sociale, il atteignait un record historique."
Faiblesse du dollar	hr_2	0.788	"Le Yen faisait toujours preuve de fermeté face au billet vert. De même, l'Euro a encore gagné du terrain face au billet vert grâce aux perspectives encourageantes de reprise économique."
Faiblesse du dollar	mr_1	0.182	"Avec l'aide de leurs rabatteurs déguenillés, elles pesaient l'offre et la demande. Elles savaient précisément quel jour du mois les entreprises chercheraient de la monnaie locale pour payer leurs salariés."
Faiblesse du dollar	mr_2	0.261	"La révision par la Commission des prévisions pour l'année confirme en tout cas qu'il y a bien un problème. Cette année, la croissance de la zone euro ne devrait être que légèrement au-dessus de 2
Faiblesse du dollar	ur_1	-0.048	"Le football est un sport collectif qui oppose deux équipes de onze joueurs dans un stade, dont le but est de mettre un ballon sphérique dans le but adverse. Le football est sans doute le sport le plus médiatisé dans le monde."

Faiblesse du dollar	ur_2	0.007	"Le skippeur rend hommage aux concepteurs. Le pari de la légèreté et de la simplicité était osé. Les gens qui se sont investis pour que ce projet existe peuvent être fiers."
Faiblesse du dollar	ur_3	-0.008	"Si la Californie avait appliqué sa politique de réduction des émissions de gaz, près de la moitié des véhicules vendus aux États-Unis auraient pu être concernés à moyen terme."
Formation en informatique	hr_1	0.719	"La plupart des animateurs-modérateurs ont une formation de journaliste. Il n'existe pas de formation spécifique, mais des écoles de journalisme ou de communication qui donnent des cours en multimédia."
Formation en informatique	hr_2	0.644	"Une bonne connaissance du fonctionnement de l'ordinateur et d'Internet est importante. Des systèmes de formation à distance et de la pédagogie pour adultes en formation continue vont être mis en place."
Formation en informatique	mr_1	0.095	"Bien plus que la connaissance d'une langue morte, l'apprentissage du latin dès le collège est un exercice du raisonnement et d'affirmation du caractère. Il est un puissant signe de différence et de condition supérieure."
Formation en informatique	mr_2	0.136	"Le Conseil national des programmes et les observatoires académiques des pratiques pédagogiques seront chargés de définir les compétences communes dispensées au collège. Les enseignants souhaitent être consultés sur ce programme."
Formation en informatique	ur_1	-0.056	"Chez les démocrates, les électeurs débattent de leurs choix respectifs et tentent de persuader leurs voisins de se rallier à leur candidat. Côté républicain, le scrutin se déroule rapidement après l'ouverture des assemblées."
Formation en informatique	ur_2	-0.011	"Facile à cultiver, décorative, variée et tellement savoureuse quand on la cueille à même le pied, la tomate est incontournable au potager. Venue d'Amérique centrale, il faut lui fournir un sol riche et une bonne dose de soleil."
Formation en informatique	ur_3	0.003	"Un domaine de réflexion a été identifié. La réticence à commercialiser en ligne est de plus en plus grande chez certains propriétaires de contenus par crainte du piratage et du téléchargement illégal."
Grève des cheminots	hr_1	0.717	"Cette grève et cette manifestation affichaient des positions et des objectifs contradictoires. La direction croyait que le mouvement allait pourrir. Elle a eu tort de spéculer sur un recul des salariés."

Grève des cheminots	hr_2	0.539	"Les transports connaissaient toujours d'importantes perturbations au quatrième jour d'une grève lancée par les syndicats pour protester contre le projet de réduction du temps de travail présenté par la direction. "
Grève des cheminots	mr_1	0.281	"La direction de l'entreprise confirme l'existence du document incriminé jugeant que l'initiative d'un collaborateur externe au processus de recrutement est inadmissible. Elle ne cautionne absolument pas cette action."
Grève des cheminots	mr_2	0.258	"Le malaise des personnels porte également sur une réforme de leur statut. Ils ont été avertis par la direction de la teneur de cette réforme juste avant la réunion du comité qui devait procéder à son examen."
Grève des cheminots	ur_1	0.029	"La Commission européenne annonce qu'elle fera des propositions pour encourager un marché commun du secteur de la musique, des films et des jeux sur internet. La situation doit être maitrisée."
Grève des cheminots	ur_2	0.015	"Le 3e ligne du XV d'Angleterre aux 85 sélections a annoncé sa retraite internationale dans les colonnes du tabloïd. Une retraite bien méritée avec une vie bien menée."
Grève des cheminots	ur_3	0.010	"Pour cette grande fresque historique, les scénaristes ont choisi de mêler étroitement l'histoire du château et celle du monarque. Celui-ci avait conçu Versailles comme l'instrument de son pouvoir absolu."
Hausse de la bourse	hr_1	0.878	"La bourse de Tokyo a clôturé la séance de mardi en légère hausse, soutenue par les achats des investisseurs étrangers. L'indice Nikkei affichait un gain important à la clôture."
Hausse de la bourse	hr_2	0.795	"L'indice vedette de la Bourse de Londres a clôturé mardi, en hausse. Il était encouragé par les diverses annonces de sociétés et nourri par les spéculations de rapprochements industriels."
Hausse de la bourse	mr_1	0.141	"Cette annonce, qui avait déjà été pronostiquée par certains analystes la semaine dernière, a été fraîchement accueillie. Cette diminution ne fait que poursuivre une tendance à la baisse du titre."
Hausse de la bourse	mr_2	0.243	"Cette réaction conduit industriels et milieux financiers à la prudence. D'autant que le champion des OGM a vu son chiffre d'affaire baisser. Les valeurs de biotechnologie ont perdu de leur attrait."

Hausse de la bourse	ur_1	0.065	"Il faut distinguer culture et civilisation. La culture est l'ensemble des croyances, des valeurs propres à une communauté particulière. La civilisation, c'est ce qui est transmis d'une communauté à une autre."
Hausse de la bourse	ur_2	0.214	"Le football américain est un sport collectif connu pour mélanger des stratégies complexes et un jeu physique intense. Le but du jeu est de marquer des points en portant un ballon ovale jusqu'à une zone de but nommée end."
Hausse de la bourse	ur_3	-0.034	"Les moines franciscains, qui ne vivent pas cloîtrés mais partagent un bâtiment moderne, ont décidé de se retrouver sur la place. Ils protestent contre les conditions de détention au centre de rétention."
Observation des planètes	hr_1	0.585	"L'étude de la dynamique des anneaux qui entourent toutes les planètes géantes en orbite autour du Soleil aide les astronomes à comprendre le processus de formation des lunes du système solaire."
Observation des planètes	hr_2	0.616	"L'étude du système solaire est difficile. Ecartons d'emblée les quatre planètes géantes qui contiennent de la glace dans leur noyau et de l'eau dans leur atmosphère pour nous concentrer sur les planètes dites telluriques."
Observation des planètes	mr_1	0.138	"Il y a une multitude de variantes qui vont de l'impossibilité pour la sonde de pointer son antenne vers nous, à un basculement de l'engin qui aurait été déséquilibré par un accident de terrain."
Observation des planètes	mr_2	0.214	"Une équipe d'un laboratoire japonais a mis au point un logiciel de traitement d'images capable de deviner la présence de mines à partir de photographies aériennes et d'un dispositif de filtrage."
Observation des planètes	ur_1	-0.006	"La plainte avait été déposée par l'entreprise et portait sur le mobilier urbain et le service de location de vélos en libre service. Le groupe américain a déposé un recours."
Observation des planètes	ur_2	-0.104	"Laure Manaudou a remporté la médaille d'argent sur 200 m nage libre aux championnats d'Europe de natation en petit bassin. Le titre est cependant revenu à la suédoise."
Observation des planètes	ur_3	0.068	"Le retour à la vraie vie peut s'avérer douloureux pour ceux qui n'ont pas économisé les dollars pour garnir le matelas qui atténuerait leur atterrissage. C'est le cas de ce basketteur célèbre."

Plantation des fleurs	hr_1	.374	"Il aura suffi que le soleil montre le bout de son nez pour que les jardineries soient prises d'assaut. On dépense de plus en plus pour la plantation de fleurs dans son jardin."
Plantation des fleurs	hr_2	.302	"Pour obtenir de belles roses, le jardinier plante les rosiers dès l'automne de façon à ce que leurs racines aient le temps de s'ancrer dans un sol encore chaud."
Plantation des fleurs	mr_1	.173	"Il pleut sur la Haute-Saône, il fait froid, les jours sont encore courts. A part quelques forsythias héroïques, les arbustes sont nus : le printemps rase les murs."
Plantation des fleurs	mr_2	.190	"C'est aux modistes parisiennes que le printemps fait ses premières confidences. Il leur offre ses premiers bouquets, ses pailles, ses rubans, ses petits canotiers car il ne saurait y avoir de printemps sans canotier."
Plantation des fleurs	ur_1	-.035	"Une inquiétude particulière pointe quant à la responsabilité pénale des chefs d'établissement mise en jeu lors d'accidents survenus à cause d'équipements défectueux, alors que l'entretien en incombe aux collectivités locales."
Plantation des fleurs	ur_2	-.069	"Trois jours après le tir du missile indien Agni-2, le Pakistan a répliqé en tirant une version améliorée de son missile balistique de moyenne portée, le Ghauri, depuis le centre d'essai de Jhelum."
Plantation des fleurs	ur_3	.058	"Le médecin en chef du service de santé des armées est rentré, il y a quelques semaines, de Terre Adélie où il a passé plus d'un an en tant que directeur et responsable médical."
Prix du pétrole	hr_1	0.587	"L'économie algérienne dépend de l'évolution des cours du pétrole brut. Après être tombés à des niveaux jamais observés, l'annonce par les pays producteurs d'une réduction de la production a redressé les prix."
Prix du pétrole	hr_2	0.598	"Le ministre irakien du pétrole a accusé dimanche l'Arabie saoudite d'avoir inondé le marché du pétrole. Il a réclamé une réduction de la production saoudienne pour que les prix remontent."
Prix du pétrole	mr_1	0.145	"L'idée est de pousser les compagnies pétrolières à réorienter leurs investissements vers la zone la moins chère de production. Dans cette région, le coût d'exploitation des gisements est très oscillant."

Prix du pétrole	mr_2	0.194	"Outre la pollution de surface, se pose aussi le problème des tonnes de pétrole encore contenues dans les soutes du navire. Les sondages effectués indiquent que les morceaux du bateau sont stabilisés."
Prix du pétrole	ur_1	-0.001	"Comment la France pourra-t-elle se passer de la cigarette ? L'interdiction de fumer dans les lieux publics est largement perçue par la presse étrangère comme une atteinte à la culture française."
Prix du pétrole	ur_2	0.037	"La Californie a officiellement déposé plainte contre le gouvernement des États-Unis pour avoir empêché l'application de sa loi de réduction des émissions de gaz à effet de serre des automobiles."
Prix du pétrole	ur_3	-0.035	"Le vieux moine est allé lui-même faire de nouveaux clichés sur place pour les publier quelques heures plus tard sur le site Internet de la communauté. Ceux-ci ont eu un succès immédiat."
Prix littéraire	hr_1	.321	"L'écrivain Amélie Nothomb présidera le jury du Livre Inter. Les auditeurs qui souhaitent participer au jury sont invités à écrire à la station en exposant leurs goûts littéraires."
Prix littéraire	hr_2	.593	"Destiné à récompenser une oeuvre de fiction inédite en prose, écrite par un auteur de moins de vingt-cinq ans, le quinzième prix du jeune écrivain a été décerné à Bernard Magnan."
Prix littéraire	mr_1	.112	"Le jeu favori des biographes de Conrad, le célèbre écrivain anglais, consiste à mesurer l'écart entre les événements de sa vie réelle et les romans qu'il en a tirés."
Prix littéraire	mr_2	.109	"Le 6e festival de Gérardmer, consacré au genre fantastique, présente une quinzaine de longs métrages inédits. En ouverture, hors compétition, Psycho, de Gus van Sant, revisite le classique d'Hitchcock."
Prix littéraire	ur_1	-.009	"Toutes hélices dehors, bigarré comme aucun avion occidental n'oserait plus l'être aujourd'hui, le cargo militaire Antonov est la vedette du salon du Bourget. Il se veut le concurrent de l'Hercules C-130J américain."
Prix littéraire	ur_2	.009	"La voiture de la présidente de la cour d'assises de Haute-Corse a été incendiée dans la nuit de samedi à dimanche, à Bastia. Selon les enquêteurs, il s'agirait d'un acte criminel."
Prix littéraire	ur_3	.022	"La crue du Gapeau, déclenchée par vingt-quatre heures de fortes pluies, a inondé trois cents maisons à Hyères et près de deux cents personnes ont dû être mises en sécurité."

Réchauffement climatique	hr_1	0.579	"L'impact du tourisme sur l'évolution du climat pourrait plus que doubler dans les trente prochaines années en raison de l'effet de serre produit par les rejets massifs de carbone associés aux activités touristiques."
Réchauffement climatique	hr_2	0.830	"Quand nous brûlons des combustibles fossiles, nous émettons des gaz à effet de serre dans l'atmosphère, ce qui a pour conséquence de la réchauffer et de contribuer à la fonte des glaciers."
Réchauffement climatique	mr_1	0.125	"Les pluies intenses ont duré plusieurs heures, causant l'interruption de la circulation et isolant un certain nombre de régions. Le bureau prévoit une augmentation de ces dérèglements climatiques au niveau de la planète."
Réchauffement climatique	mr_2	0.104	"Un groupe d'ingénieurs météorologistes a obtenu le prix Nobel pour ses travaux relatifs à la prévision des évènements climatiques à long terme et aux effets de la pollution de l'atmosphère."
Réchauffement climatique	ur_1	0.013	"Dans le jugement, cité par La Nacion, les juges mettent en avant plusieurs arguments de forme, arguant que le type de procédure lancé par le juge Cerda serait inapproprié."
Réchauffement climatique	ur_2	-0.001	"La Russie a menacé de sortir du traité sur les forces nucléaires car elle est courroucée par l'attitude des autres pays vis-à-vis de sa politique de défense nationale."
Réchauffement climatique	ur_3	-0.039	"Une famille s'approche lentement du juge. Le père tient sa petite fille dans les bras. La mère pose son sac sur le bureau et cherche ses justificatifs."
Réforme de l'enseignement	hr_1	0.455	"La critique s'est exprimée plus radicalement sur la nature et l'idéologie du projet éducatif. Dans ce débat, déclenché en dehors des organisations syndicales, les enseignants s'en prennent aux orientations fondamentales."
Réforme de l'enseignement	hr_2	0.597	"Malgré un accord de principe sur la nécessité d'une réforme comprenant une aide individualisée aux élèves, le syndicat des enseignants affirmait qu'il fallait abandonner ce projet et retravailler sur d'autres bases."
Réforme de l'enseignement	mr_1	0.296	"Afin d'apprendre une langue étrangère le ministère entend embaucher des locuteurs natifs pour faire pratiquer l'oral aux élèves. Cette politique est généralisée pour les enseignements communs, les options obligatoires et facultatives."

Réforme de l'enseignement	mr_2	0.147	"Après avoir rappelé que les gouvernements successifs et le corps enseignant ont su répondre au défi démographique, le ministre a rappelé que l'effort devait à présent garantir la qualité pour tous."
Réforme de l'enseignement	ur_1	-0.023	"Cette saison, Tony Parker n'a plus confiance sur le terrain. Il y a des périodes pendant les matches où il comprend qu'il ne peut pas totalement dominer."
Réforme de l'enseignement	ur_2	-0.045	"Les labels alertent le public avec le même refrain : les plates-formes de téléchargement illégales tuent l'industrie musicale. Pendant ce temps, les marchands de disques ont le moral en berne."
Réforme de l'enseignement	ur_3	0.054	"Le volume de fusions-acquisitions dans le monde a atteint un record historique. Ce résultat a été obtenu malgré un fléchissement au second semestre, lié aux inquiétudes sur le marché du crédit."
Réforme de la justice	hr_1	0.757	"Le président de l'APM a vilipendé la réforme de la ministre de la justice, mais il a surtout mené la charge contre les élus, accusés de dresser le procès de la magistrature."
Réforme de la justice	hr_2	0.402	"Mentir sous serment, c'est une insulte à la liberté. Faire obstruction à la justice, c'est bafouer la loi. Il y a des gens en prison pour de tels crimes."
Réforme de la justice	mr_1	0.106	"Bien que les preuves de la discrimination à l'embauche soient souvent difficiles à établir devant les juges, quelques affaires ont pu aboutir à confondre des employeurs et ont entraîné des condamnations."
Réforme de la justice	mr_2	0.133	"Le Code de procédure pénale réaffirme le principe de la publicité des débats. Il indique que les débats sont publics, à moins que la publicité ne soit dangereuse pour l'ordre."
Réforme de la justice	ur_1	-0.079	"La science n'apporte pas seulement des bienfaits, mais a apporté des armes de destruction massive et des possibilités de manipulation biologique. La technique et l'économie concourent à la dégradation de la biosphère."
Réforme de la justice	ur_2	-0.002	"Le chanteur Yannick Noah, ici en concert, a repris sa place à la tete du classement des personnalités préférées des Français et devance Zinédine Zidane."
Réforme de la justice	ur_3	-0.019	"L'industrie musicale a enregistré un nouveau plongeon des ventes d'albums. Les ventes totales ont baissé fortement et il a fallu attendre pour voir décoller la meilleure vente de l'année."

Réhabilitation des logements	hr_1	0.402	"Aux abords de la gare, nulle barre de logements, ni cabine téléphonique dévastée, mais un habitat constitué d'immeubles réhabilités et de maisons en pierre meulière restaurées et ornées de charmants balcons."
Réhabilitation des logements	hr_2	0.448	"Un décret a été émis qui modifie le code de la construction et de l'habitation. Il est relatif aux subventions et prêts pour la construction, l'acquisition et l'amélioration des logements locatifs aidés."
Réhabilitation des logements	mr_1	0.349	"Le revenu moyen imposable par foyer fiscal est le plus faible des communes de l'agglomération. Riches-pauvres : on n'échappe pas au face-à-face. C'est celui qui embrase la plupart des convulsions de l'histoire."
Réhabilitation des logements	mr_2	0.119	"Le rythme des transformations de bureaux s'accélère dans la capitale. En réalité, la moitié deviennent des appartements intermédiaires d'un niveau plus élevé que celui du parc classique."
Réhabilitation des logements	ur_1	0.074	"La baie de Somme en hiver : l'espace, le silence, des instants fugaces et des lumières magiques. Comme ce coucher de soleil sur la pointe du Hourdel."
Réhabilitation des logements	ur_2	-0.039	"La Chine souhaite généraliser l'usage de l'injection mortelle pour les peines capitales au détriment de l'exécution par balle. C'est le premier pays au monde pour le nombre d'exécutions par an."
Réhabilitation des logements	ur_3	-0.038	"Le badminton nécessite de nombreuses heures d'entraînement par semaine, y compris chez les jeunes. Il est difficile de faire l'impasse sur une compétition puisque l'athlète n'échappe jamais à une logique de classement."
Salon aéronautique	hr_1	.510	"Un avion de combat russe biréacteur Sukhoi-30 s'est écrasé pendant une démonstration en vol, samedi vers 15 heures, lors de l'inauguration du salon aéronautique du Bourget."
Salon aéronautique	hr_2	.310	"Le représentant du consortium aéronautique européen pouvait parader dans les stands : Airbus a reçu une commande de 52 avions dès le premier jour du salon du Bourget."
Salon aéronautique	mr_1	.109	"L'exploit de Louis Blériot, le premier aviateur à avoir traversé la manche en avion en 1909, a été reproduit dimanche 25 juillet, 90 ans jour pour jour aprés, par un pilote de ligne suédois."

Salon aéronautique	mr_2	.178	"Un miraculé de l'aviation : un jeune passager clandestin a fait le trajet Dakar-Lyon, mardi dernier, dans le train d'atterrissage d'un Airbus A300.
Salon aéronautique	ur_1	.012	"Qui sont ces étudiants qui n'ont pas les moyens de s'offrir un repas à 3 euros ? Principalement des étudiants étrangers qui n'ont pas de bourses d'études, résume l'assistante sociale du Crous.
Salon aéronautique	ur_2	-.002	"La neige fait défaut dans les Hautes-Alpes. Cinq cents travailleurs saisonniers privés d'emploi ont manifesté dans les rues de Gap pour exiger des exploitants de remontées mécaniques des indemnisations."
Salon aéronautique	ur_3	-.046	"Une gardienne de prison a été condamnée à six mois de prison avec sursis par la cour d'appel de Lyon. Il lui était reproché d'avoir offert un téléphone portable à une détenue."
Syndicat agricole	hr_1	.723	"Un millier d'agriculteurs ont paralysé le centre d'Auch pour exprimer leur inquiétude sur la réforme de la politique agricole commune, à l'appel de la Fédération Départementale des Syndicats d'Exploitants Agricoles."
Syndicat agricole	hr_2	.564	"Les producteurs de fruits et légumes, surtout affiliés à la puissante Fédération Nationale des Syndicats d'Exploitants Agricoles, ont les grandes surfaces dans leur ligne de mire."
Syndicat agricole	mr_1	.146	"L'indivision est un fléau en Corse, surtout dans l'intérieur des terres. Elle empêche les jeunes agriculteurs de s'installer, faute de savoir à qui racheter la terre."
Syndicat agricole	mr_2	.223	"Les syndicats qui avaient lancé l'ordre de grève réclamaient une amélioration de l'aménagement du temps de travail, ainsi que l'embauche de trente personnes supplémentaires."
Syndicat agricole	ur_1	.003	"Le moral des boursiers s'est sérieusement dégradé : ils doutent désormais que les nouvelles marges de fluctuation de la devise brésilienne, fixées mercredi par la banque centrale, puissent tenir longtemps."
Syndicat agricole	ur_2	.027	"En Inde, l'épais brouillard, qui recouvre le nord du pays depuis plusieurs jours et qui s'ajoute à une vague de froid dans l'état du Rajasthan, paralyse les transports aériens."
Syndicat agricole	ur_3	-.029	"Le gouvernement chypriote a émis l'espoir que sa décision de renoncer à déployer dans le sud de l'île des missiles sol-air russes facilitera l'adhésion de Chypre à l'union européenne."

Tourisme en montagne	hr_1	0.422	"Les via ferrata et les sports d'eau vive connaissent un succès constant. Les candidats au vertige se comptent par milliers sur la cinquantaine de parcours de via ferrata ancrés en pleine falaise."
Tourisme en montagne	hr_2	0.382	"Toute la magie des hauts plateaux basaltiques avec la neige en plus. Au programme : randonnée en raquettes avec guide, ski de fond et relaxation dans les eaux chaudes thermales de la montagne."
Tourisme en montagne	mr_1	0.226	"Pour les vacanciers qui souhaitent avant tout se reposer dans un cadre bucolique, les résidences proposent un hébergement adapté leur permettant de partir avec des amis sportifs sans pour autant rester inactifs."
Tourisme en montagne	mr_2	0.212	"Une semaine à Santorin, l'une des plus belles et, assurément, en raison de son caractère volcanique, la plus spectaculaire des îles grecques est proposée par l'agence de voyage."
Tourisme en montagne	ur_1	0.031	"Des trafiquants de cocaïne colombiens ont été sabordés par un sous-marin au large de la côte Pacifique. C'est la deuxième fois que les forces colombiennes arrêtent un engin afin d'en détruire la cargaison illicite."
Tourisme en montagne	ur_2	-0.027	"Il faut profiter de la crise générale que traverse le gouvernement et de la fatigue de certaines unités militaires. Le guérillero juge nécessaire de mener des actions armées sur les routes."
Tourisme en montagne	ur_3	0.012	"Le chef des Forces armées révolutionnaires de Colombie demande à ses troupes de commencer à préparer les conditions d'une offensive générale contre le Président Uribe."
Tribunal international	hr_1	0.798	"Les États-Unis et de nombreuses ONG préparent juridiquement une mise en accusation, pour crimes contre l'humanité devant le Tribunal pénal international, des dirigeants civils, militaires et paramilitaires de Belgrade."
Tribunal international	hr_2	0.797	"Nos gouvernements coopéreront avec le Tribunal pénal international pour l'ex-Yougoslavie en l'aidant à enquêter sur les responsables, jusqu'aux plus hauts niveaux, de crimes de guerre et de crimes contre l'humanité."
Tribunal international	mr_1	0.136	"Principal accusé du volet corruption et escroquerie, un ancien cadre d'une grande entreprise est poursuivi pour avoir perçu des millions d'euros en commissions occultes auprès d'une dizaine d'entrepreneurs."

Tribunal international	mr_2	0.183	"Un couple de quinquagénaires mosellans sera jugé mardi en correctionnel pour détention et usage personnel de stupéfiants. Ils arrosaient leurs pots de cannabis sur le rebord de leur fenêtre."
Tribunal international	ur_1	0.038	"Aucune preuve scientifique ne permet aujourd'hui de démontrer que l'utilisation des téléphones mobiles présente un risque notable pour la santé. Le communiqué du ministère ne laisse place à aucun doute."
Tribunal international	ur_2	0.053	"La Patagonie chilienne est mal connue. Coincée entre la cordillère des Andes et l'océan Pacifique, elle séduit les amateurs de grands espaces, de solitude et de paysages grandioses."
Tribunal international	ur_3	-5.337	"Partir pour la gagne et se retrouver en panne entre deux éléphants de mer, ça laisse des souvenirs. Un tour du monde à la voile reste une sacrée aventure."
Victoire des footballeurs	hr_1	0.512	"Dans ce contexte relevé, les joueurs marseillais voulaient surtout satisfaire leur public. Avec cette finale gagnée, l'équipe phocéenne a particulièrement réussi son retour parmi l'élite du football professionnel."
Victoire des footballeurs	hr_2	0.490	"L'équipe de France de football a vaincu l'Australie en finale de la Coupe du Monde. L'entraineur est très satisfait de son équipe et envisage les rencontres futures avec enthousiasme et sérénité."
Victoire des footballeurs	mr_1	0.228	"Il a inventé une forme de tennis qui lui convient parfaitement et le rend toujours dangereux. Avec cette façon d'engager des revers, sa raquette réalise des coups souvent surprenants."
Victoire des footballeurs	mr_2	0.259	"Il semble que, à l'instar des Anglais, les Gallois soient parvenus à assimiler les éléments qui ont permis à l'hémisphère Sud de dominer le rugby mondial depuis quelques années."
Victoire des footballeurs	ur_1	0.032	"Des livres de mémoires d'anciens déportés, il en paraît régulièrement. De qualité inégale et d'intérêt inégal mais ce n'est jamais la même histoire car le ressenti est à chaque fois unique."
Victoire des footballeurs	ur_2	0.093	"Le candidat conservateur a bénéficié du soutien du fort contingent d'évangélistes et de religieux conservateurs qui habitent cet État souvent critiqué pour son manque de représentativité à l'échelon national."
Victoire des footballeurs	ur_3	-0.057	"Les rebelles ivoiriens toujours maîtres de la moitié nord du pays même s'ils participent au gouvernement ont reconnu que des troubles avaient eu lieu à Bouaké."

Bibliography

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychol Rev*, 111(4):1036–1060. (Cited on page 3.)

Baccino, T., Bellino, C., and Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces.* Hermès Science Publisher: Paris. (Cited on page 119.)

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436. (Cited on pages 48, 59 and 95.)

Browne, G. J., Pitts, M. G., and Wetherbe, J. C. (2005). Stopping rule use during web-based search. In *38th Hawaii International Conference on System Sciences (HICSS-38 2005), CD-ROM / Abstracts Proceedings, 3-6 January 2005, Big Island, HI, USA*. IEEE Computer Society. (Cited on page 31.)

Browne, G. J., Pitts, M. G., and Wetherbe, J. C. (2007). Cognitive stopping rules for terminating information search in online tasks. *MIS Q.*, 31(1):89–104. (Cited on pages 30, 31 and 32.)

Campbell, D. J. (1988). Task complexity: A review and analysis. *The Academy of Management Review*, 13(1):40–52. (Cited on page 32.)

Carver, R. (1990). *Reading rate: a review of research and theory.* Academic Press. (Cited on pages 7, 8, 26, 98 and 118.)

Chanceaux, M. (2009). *Modélisation cognitive computationnelle de trajets oculomoteurs lors d'une tâche de recherche d'information.* PhD thesis, Laboratoire TIMC-IMAG, équipe AMA. Université de Grenoble. (Cited on pages 28 and 29.)

Chanceaux, M., Guérin-Dugué, A., Lemaire, B., and Baccino, T. (2009). A model to simulate web users' eye movements. In Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R., and Winckler, M., editors, *Proceedings of the 12th INTERACT Conference*, volume 5726 of *LNCS*, pages 288–300. Springer Verlag. (Cited on pages 92 and 119.)

Cole, M. J., Gwizdka, J., Liu, C., Bierig, R., Belkin, N. J., and Zhang, X. (2011). Task and user effects on reading patterns in information search. *Interacting with Computers. Cognitive Ergonomics for Situated Human-Automation Collaboration*, 23(4):346–362. (Cited on page 26.)

Engbert, R., Longtin, A., and Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636. (Cited on page 18.)

Engbert, R., Nuthmann, A., Richter, E., and Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4):777–813. (Cited on pages 16, 18 and 20.)

Farid, M. and Grainger, J. (1996). How initial fixation position influences word recognition: A comparison of french and arabic. *Brain & Language*, 53(3):351–368. (Cited on page 51.)

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In Sun, R., editor, *Cambridge handbook of computational cognitive modeling*, pages 59–100. Cambridge University Press, Cambridge. (Cited on page 53.)

Griffiths, T. L. and Steyvers, M. (2002). A probabilistic approach to semantic representation. (Cited on page 37.)

Hofmann, T. (1999). Probabilistic Latent Semantic Analysis. In *Proceedings of Uncertainty in Artificial Intelligence, UAI*, Stockholm. (Cited on page 118.)

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506. (Cited on page 28.)

Kieras, D. and Meyer, D. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. (Cited on page 3.)

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in psychtoolbox-3? In *Perception 36 ECVP Abstract Supplement*. Perception 36 ECVP Abstract Supplement. (Cited on pages 48, 59 and 95.)

Kohsom, C. and Gobet, F. (1997). Adding spaces to thai and english: Effects on reading. In *Proceedings of the Cognitive Science Society*, volume 19, pages 388–393. Cognitive Science Society. (Cited on page 15.)

Landauer, T., McNamara, D., Dennis, S., and Kintsch, W. (2007). *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates. (Cited on pages 29, 37, 41, 47, 52, 61, 103, 118, 151 and 152.)

Landauer, T. K. and Dutnais, S. T. (1997). A solution to plato‘s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104:211–240. (Cited on pages 52, 61 and 103.)

Lee, M. D. and Corlett, E. Y. (2003). Sequential sampling models of human text classification. *Cognitive Science*, 27(2):159 – 193. (Cited on pages 33, 34, 35, 36, 37 and 57.)

Lemaire, B., Guérin-Dugué, A., Baccino, T., Chanceaux, M., and Pasqualotti, L. (2011). A cognitive computational model of eye movements investigating visual strategies on textual material. In Carlson, L., Hölscher, C., and Shipley, T., editors, *Proceedings of the Annual Meeting of the Cognitive Science Society*, pages 1146–1151. Cognitive Science Society. (Cited on pages 8, 26, 27 and 118.)

Lewis, D. D. (1997). Reuters-21578 text categorization test collection. (Cited on page 34.)

Lund, K. and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2):203–208. (Cited on page 37.)

Martin, D. I. and Berry, M. W. (2007). Mathematical Foundations Behind Latent Semantic Analysis. In Landauer, T. K., McNamara, D. S., Dennis, S., and Kintsch, W., editors, *Handbook of Latent Semantic Analysis*, chapter 2. Lawrence Erlbaum Associates. (Cited on page 42.)

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychol Bull*, 81(12):899–917. (Cited on page 9.)

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press, Cambridge, MA, USA. (Cited on pages 2 and 3.)

Nickles, K. (1995). *Judgment-based and Reasoning-based Stopping Rules in Decision Making Under Uncertainty.* University of Minnesota. (Cited on page 31.)

Nuthmann, A., Engbert, R., and Kliegl, R. (2007). The IOVP effect in mindless reading: experiment and modeling. *Vision Res*, 47(7):990–1002. (Cited on pages 11 and 12.)

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442. (Cited on pages 48, 59 and 95.)

Pollatsek, A., Raney, G. E., LaGasse, L., and Rayner, K. (1993). The use of information below fixation in reading and in visual search. *Canadian Journal of Experimental Psychology*, 47:179–200. (Cited on page 13.)

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1):65–81. (Cited on page 14.)

Rayner, K. (1978). Eye movement latencies for parafoveally presented words. *Bulletin of the Psychonomic Society*, 11(1eye movement latencies, parafoveally presented words; right handed adults):13–16+. (Cited on page 9.)

Rayner, K. (1979). Eye guidance in reading: Fixation locations in words. *Perception*, 8:21–30. (Cited on pages 14 and 15.)

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422. (Cited on pages 9, 10, 13, 15, 24 and 51.)

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8):1457–1506. (Cited on pages 9 and 15.)

Rayner, K. and Fischer, M. H. (1996). Mindless reading revisited: eye movements during reading and scanning are different. *Perception and Psychophysics*, 58(5global/local eye movements; task, word frequency; reading/scanning/visual search tasks):734–747+. (Cited on page 11.)

Rayner, K. and Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, 33A:351–373. (Cited on page 14.)

Rayner, K. and Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs NJ: Prentice-Hall. (Cited on page 10.)

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., and Reichle, E. D. (2006). Tracking the mind during reading via eye movements: Comments on kliegl, nuthmann, and engbert (2006). *Journal of Experimental Psychology: General*, 136:520–529. (Cited on page 14.)

Rayner, K., Sereno, S. C., and Raney, G. E. (1996). Eye movement control in reading: comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22:1188–1200. (Cited on page 15.)

Rayner, K., Slowiaczek, M. L., Clifton, C., and Bertera, J. H. (1983). Latency of sequential eye movements: implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, 9(6latency of saccadic eye movements, reading):912–922+. (Cited on page 9.)

Rayner, K., Well, A. D., and Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception and Psychophysics*, 27:537–544+. (Cited on page 13.)

Rayner, K., Well, A. D., Pollatsek, A., and Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, 31(6):537–550+. (Cited on page 13.)

Reichle, E., Pollatsek, A., Fisher, D., and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1):125–157. (Cited on pages 16 and 17.)

Reichle, E. D., Pollatsek, A., and Rayner, K. (2006). E-z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7(1):4–22. (Cited on pages 16 and 17.)

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The e-z reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(4):445–526. (Cited on pages 16 and 17.)

Reichle, E. D., Rayner, K., and Pollatsek, A. (2012). Eye movements in reading versus nonreading tasks: Using e-z reader to understand the role of word/stimulus familiarity. *Visual Cognition*, 20(4-5):360–390. (Cited on pages 18, 21 and 22.)

Reilly, R. G. and Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7(1):34–55. (Cited on page 16.)

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471. (Cited on page 3.)

Sartre, J. (1984). *Existentialism and Human Emotions*. New York: Carol Publishing Group. (Cited on page 29.)

Sereno, S. and Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cogntive Sciences*, 7(11):489–493. (Cited on pages 7, 10 and 11.)

Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat Neurosci*, 6(12):1317–1322. (Cited on page 101.)

Simola, J., Salojärvi, J., and Kojoc, I. (2008). Using hidden markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4):237–251. (Cited on pages 8, 23, 24, 25 and 26.)

Simon, H. A. (1996). *The sciences of the artificial (3rd ed.)*. MIT Press, Cambridge, MA, USA. (Cited on page 30.)

Sun, R. (2008). *Introduction to computational cognitive modeling.*, chapter 1, pages 3–19. New York : Cambridge University Press. (Cited on pages 1 and 2.)

Sun, R., Coward, L. A., and Zenzen, M. J. (2005). On levels of cognitive modeling. *Philosophical Psychology*, 18:613–637. (Cited on page 2.)

Underwood, N. and McConkie, G. (1983). *Perceptual Span for Letter Distinctions During Reading.* Technical report (University of Illinois at Urbana-Champaign. Center for the Study of Reading). University of Illinois at Urbana-Champaign. (Cited on page 13.)

Zandt, T., Colonius, H., and Proctor, R. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, 7:208–256. (Cited on page 35.)

Résumé: Cette thèse en informatique présente un travail de modélisation cognitive computationnelle, à partir de données de mouvements oculaires lors de tâches de recherche d'information dans des textes. Nous nous intéressons à cette situation quotidienne de recherche d'informations dans un journal ou une page web, dans laquelle il faut juger si un texte est sémantiquement relié ou non à un but, exprimé par quelques mots. Parce que le temps est souvent une contrainte, les textes ne sont souvent pas entièrement lus avant qu'intervienne la décision. Plus précisément, nous avons analysé les mouvements des yeux dans deux tâches de recherche d'information consistant à lire un paragraphe et à décider rapidement i) s'il est associé à un but donné et ii) s'il est plus associé à un but donné qu'un autre paragraphe traité auparavant. Un modèle est proposé pour chacune de ces situations. Nos simulations sont réalisées au niveau des fixations et des saccades oculaires.

En particulier, nous prédisons le moment auquel les participants décident d'abandonner la lecture du paragraphe parce qu'ils ont suffisamment d'information pour prendre leur décision. Les modèles font ces prédictions par rapport aux mots qui sont susceptibles d'être traités avant que le paragraphe soit abandonné. Les jugements d'association sémantiques humains sont reproduits par le calcul des similarités sémantiques entre mots produits par l'analyse de la sémantique latente (LSA) (Landauer et al., 2007). Nous avons suivi une approche statistique paramétrique dans la construction de nos modèles. Ils sont basés sur un classifieur bayésien.

Nous proposons un seuil linéaire bi-dimensionnel pour rendre compte de la décision d'arrêter de lire un paragraphe, utilisant le *Rang* de la fixation et i) la similarité sémantique (*Cos*) entre le paragraphe et le but ainsi que ii) la différence de similarité sémantique (*Gap*) entre chaque paragraphe et le but. Pour chacun des modèles, les performances montrent que nous sommes capables de reproduire en moyenne le comportement des participants face aux tâches de recherche d'information étudiées durant cette thèse.

Cette thèse comprend deux parties principales : 1) la conception et la passation d'expériences psychophysiques pour acquérir des données de mouvements oculaires et 2) le développement et le test de modèles cognitifs computationnels.

Mots clés: modélisation cognitive computationnelle, recherche d'information, mouvements des yeux.

Abstract: This computer science thesis presents a computational cognitive modeling work using eye movements of people faced to different information search tasks on textual material. We studied situations of everyday life when people are seeking information on a newspaper or a web page. People should judge whether a piece of text is semantically related or not to a goal expressed by a few words. Because quite often time is a constraint, texts may not be entirely processed before the decision occurs. More specifically, we analyzed eye movements during two information search tasks: reading a paragraph with the task of quickly deciding i) if it is related or not to a given goal and ii) whether it is better related to a given goal than another paragraph processed previously. One model is proposed for each of these situations. Our simulations are done at the level of eye fixations and saccades.

In particular, we predicted the time at which participants would decide to stop reading a paragraph because they have enough information to make their decision. The models make predictions at the level of words that are likely to be fixated before a paragraph is abandoned. Human semantic judgments are mimicked by computing the semantic similarities between sets of words using Latent Semantic Analysis (LSA) (Landauer et al., 2007). We followed a statistical parametric approach in the construction of our models. The models are based on a Bayesian classifier.

We proposed a two-variable linear threshold to account for the decision to stop reading a paragraph, based on the *Rank* of the fixation and i) the semantic similarity (*Cos*) between the paragraph and the goal and ii) the difference of semantic similarities (*Gap*) between each paragraph and the goal. For both models, the performance results showed that we are able to replicate in average people's behavior faced to the information search tasks studied along the thesis.

The thesis includes two main parts: 1) designing and carrying out psychophysical experiments in order to acquire eye movement data and 2) developing and testing the computational cognitive models.

Keywords: computational cognitive modeling, information search, eye movements.