



HAL
open science

Modélisation cognitive computationnelle de trajets oculomoteurs lors d'une tâche de recherche d'information

Myriam Chanceaux

► **To cite this version:**

Myriam Chanceaux. Modélisation cognitive computationnelle de trajets oculomoteurs lors d'une tâche de recherche d'information. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT: . tel-00430624v2

HAL Id: tel-00430624

<https://theses.hal.science/tel-00430624v2>

Submitted on 6 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse

pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER DE GRENOBLE

Spécialité : Informatique

Préparée au laboratoire TIMC-IMAG UMR 5525

École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE)

présentée et soutenue par

Myriam Chanceaux

le 24 novembre 2009

Modélisation cognitive computationnelle de trajets oculomoteurs lors d'une tâche de recherche d'information

Directeurs de Thèse :

Anne Guérin-Dugué et Benoît Lemaire

Composition du jury :

M	Jean Caelen (LIG, Grenoble)	Président
Mme	Mireille Bétrancourt (TECFA, Genève)	Rapporteuse
Mme	Karine Doré-Mazars (LPNCOG, Paris)	Rapporteuse
M	Frédéric Alexandre (LORIA, Nancy)	Examineur
M	Thierry Baccino (LUTIN, Paris)	Examineur
Mme	Anne Guérin-Dugué (GIPSA, Grenoble)	Directrice de Thèse
M	Benoît Lemaire (TIMC, Grenoble)	Directeur de Thèse

Résumé

Ce travail de thèse étudie la combinaison des processus visuels et sémantiques dans le traitement d'interfaces textuelles, comme des pages web. La méthodologie employée est la simulation de modèles cognitifs, une approche visant à concevoir un programme informatique s'inspirant de modèles cognitifs théoriques et reproduisant au mieux le comportement humain. Notre modèle simule les trajets oculomoteurs d'un utilisateur moyen lors d'une recherche d'information. Les processus mis en jeu dans ce type de tâche sont modélisés afin de reproduire les mouvements oculaires humains enregistrés lors de différentes expérimentations. Aux modèles des processus visuels et sémantiques est ajouté un modèle des processus mnésiques sous-jacents à la recherche d'information.

Pour la partie visuelle, le modèle s'inspire des cartes de saillance qui prédisent les zones de l'écran susceptibles d'attirer l'attention, en fonction des informations à bas niveau (couleur, orientation et contraste), ainsi que des propriétés physiologiques de l'œil humain. Pour la partie sémantique, la technique de mesure de similarités sémantiques entre le but de l'utilisateur et les différentes parties de la page employée est LSA (Latent Semantic Analysis) (Landauer, Foltz, & Laham, 1998). Pour la partie mnésique, le mécanisme d'inhibition de retour (Klein & MacInnes, 1999) et le Variable Memory Model (Horowitz, 2006) sont utilisés.

Le travail de thèse intègre trois parties : conception d'un modèle théorique d'interaction, conception de l'outil de simulation, et mise en place d'expérimentations psychophysiques de relevés oculométriques permettant de valider et d'affiner le modèle proposé.

Abstract

This thesis examines the combination of visual and semantic processes in information seeking task on textual interfaces such as web pages. The methodology used is the simulation of cognitive models. This approach aims to design a program based on theoretical cognitive models replicating human behaviour. Our model simulates the oculomotor scan path of an average user searching for information. The processes involved in this kind of tasks are modelled to replicate human eye movements recorded during different experiments. Models of visual and semantics processes are added to a model of memory processes inherent in information retrieval task.

For the visual part, we go 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. For the semantic part, the method used to measure semantic similarities between the search goal of the user and the different parts of the page is LSA (Latent Semantic Analysis) (Landauer et al., 1998). For the memory part the mechanism of inhibition of return (Klein & MacInnes, 1999) and the Variable Memory Model (Horowitz, 2006) are used.

The thesis includes three parts : designing a theoretical model of interaction, designing a simulation tool, and developing psychophysical experiments with eye-tracking techniques to validate and refine the proposed model.

Remerciements

Je tiens ici à remercier toutes les personnes sans lesquelles ce travail n'aurait pas été celui que je présente aujourd'hui.

Ces trois années ont été possibles et ont vu le jour grâce à Anne et Benoit qui m'ont parfaitement guidée dans la voie de la recherche, me conseillant et me corrigeant tout en me laissant libre dans mes choix. Ils ont été très présents tout au long de cette thèse et je les remercie pour leurs qualités à la fois scientifiques et personnelles.

Merci ensuite aux membres de mon jury pour avoir consacré une partie de leur temps à l'évaluation de ces travaux. Je remercie Jean Caelen pour avoir accepté de présider ce jury. Merci également à mes rapporteuses, Mireille Bétrancourt et Karine Doré-Mazars, pour leur relecture attentive de mon manuscrit et leurs rapports détaillés et pertinents. Merci à Frédéric Alexandre, qui a accepté d'évaluer ce travail, et Thierry Baccino pour sa participation à mon jury mais aussi et surtout pour ces trois années de collaboration riches et fructueuses. Je tiens aussi à remercier John Henderson pour son accueil à Edinburgh, ainsi que tous les membres du *Visual Cognition Lab* où j'ai travaillé durant 6 mois.

De manière plus générale, merci aux membres de l'équipe AMA qui en font une équipe très agréable et stimulante, et grâce auxquels j'ai eu plaisir à venir travailler ces dernières années.

Je finirai par des remerciements plus personnels qui iront à Manu, à ma maman, à la famille et aux amis.

Table des matières

Introduction	1
1 Modélisation cognitive de la recherche d'information	9
1.1 Guidage ascendant de l'attention	10
1.1.1 Théorie d'intégration des traits	11
1.1.2 Modèle de cartes de saillance de Itti	12
1.2 Modèles de recherche visuelle	13
1.2.1 Modèle <i>Guided Search</i>	14
1.2.2 Modèle <i>Area Activation</i>	15
1.2.3 Modèle de Navalpakkam et Itti	15
1.3 Modèles de recherche dans des documents textuels	17
1.3.1 Modèles de lecture	17
1.3.2 SNIF-ACT	23
1.3.3 CoLiDeS	26
1.3.4 Complémentarité de CoLiDeS et SNIF-ACT	28
1.4 Problématique	28
2 Présentation générale du modèle	31
2.1 Application du modèle	32
2.2 Architecture	34
2.3 Processus visuel	36
2.3.1 Éléments théoriques	36

2.3.2	Pistes de modélisation	39
2.4	Processus sémantique	39
2.4.1	Éléments théoriques	39
2.4.2	Analyse de la Sémantique Latente, LSA	40
2.4.3	Piste de modélisation	50
2.5	Processus mnésique	50
2.5.1	Éléments théoriques	50
2.6	Modélisation	53
2.6.1	Processus visuel	53
2.6.2	Processus sémantique	57
2.6.3	Processus mnésique	60
2.6.4	Discussion	62
3	Application du modèle à un environnement simple	65
3.1	Expérimentations avec des mots	66
3.1.1	Méthode	66
3.1.2	Déroulement de l'expérience	72
3.1.3	Résultats expérimentaux	73
3.2	Modélisation et comparaisons avec les données empiriques	80
3.2.1	Les trois processus et leur interaction	80
3.2.2	Simulations	83
3.2.3	Comparaison des données	83
3.3	Discussion	86
4	Environnement complexe : matériel expérimental	91
4.1	Choix des textes	92
4.2	Organisation en lignes et colonnes	93
4.2.1	Information visuelle : photo	93
4.2.2	Information visuelle : blocs en couleur	94
4.2.3	Contraintes de placement	94
4.3	Évaluation du matériel	95
4.3.1	Test de la photo	95
4.3.2	Modification du matériel	96
4.4	Organisation en pêle-mêle	97
5	Environnement complexe : parcours inter-blocs	99
5.1	Modélisation du parcours inter-blocs	99
5.1.1	Premières visites des blocs	100
5.1.2	Revisites	101
5.2	Expérimentations	102
5.2.1	Dispositif	102
5.2.2	Participants	104
5.3	Ajustement du modèle	104

5.4	Comparaison du modèle avec les données empiriques	108
5.4.1	Transitions entre les blocs	108
5.4.2	Revisites des blocs	110
5.5	Discussion	112
6	Environnement complexe : parcours intra-bloc	113
6.1	Résultats expérimentaux	114
6.1.1	Lecture vs exploration	114
6.1.2	Traitement du titre	116
6.2	Modélisation	118
6.2.1	Analyse morpho-syntaxique	118
6.2.2	Similarité entre le texte et le thème	119
6.2.3	Prise en compte du titre	120
6.2.4	Exemple de trajet du modèle	124
6.3	Comparaison du modèle avec les données empiriques	126
6.3.1	Taux de lecture	126
6.3.2	Utilisation du titre	128
6.4	Discussion	129
	Conclusion et perspectives	133
	Discussion générale	133
	Perspectives de recherche	136
	Annexes	137
A	Mots utilisés pour les expériences	137
B	Questionnaire	142
C	Matrice de similarités	145
D	Pages pour les différentes questions de l'expérience avec les mots	147
E	Pages pour les différentes modalités de l'expérience avec les mots	151
F	Textes pour l'expérience avec les paragraphes	154
G	Pages d'un thème pour l'expérience structurée en lignes et colonnes	161
H	Pages d'une liste pour l'expérience structurée en lignes et colonnes	163
I	Pages d'un thème dans la modalité <i>Sans Titre</i> pour l'expérience structurée en pêle-mêle	168
J	Pages d'une liste pour l'expérience structurée en pêle-mêle	170
K	Pages d'une liste dans la modalité <i>Avec Titre</i> pour l'expérience structurée en pêle-mêle	175
	Références	181

Introduction

Cette thèse d'informatique se situe dans le champ de recherche des sciences cognitives, et plus particulièrement dans celui de la modélisation cognitive computationnelle. Avec cet objectif de simulation sur ordinateur des processus cognitifs, nous allierons différents domaines de recherche : l'informatique, se rapprochant de l'Intelligence Artificielle, et la psychologie cognitive et expérimentale pour ce qui relève des processus humains et de la validation par des expérimentations.

Modéliser pour comprendre

Un modèle cognitif computationnel est tout d'abord un modèle cognitif, c'est-à-dire une représentation simplifiée des processus cognitifs et psychologiques permettant de décrire le comportement humain. Les modèles cognitifs peuvent se séparer en trois classes : les modèles théoriques, les modèles mathématiques et les modèles computationnels (Sun, 2008).

Les modèles théoriques sont « sur papier » et mettent en relation des concepts cognitifs afin de décrire ou expliquer les comportements. Pour tester et valider ces modèles le cycle classiquement utilisé est *Modèle* \leftrightarrow *Hypothèses* \leftrightarrow *Expérimentation*, représenté figure 1 page suivante. À partir d'un modèle du fonctionnement cognitif, des hypothèses sont formulées puis vérifiées (ou non) par des expérimentations. De même, à l'issue d'observations comportementales des hypothèses de fonctionnement sont formulées qui conduisent à une modélisation qui pourra fournir d'autres hypothèses, etc. Prenons par exemple un modèle de la mémoire qui suppose deux entités

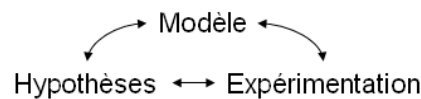


FIGURE 1 – Cycle de modélisation

séparées : mémoire à court terme et à long terme, une hypothèse issue de ce modèle est la suivante : un court-circuit dans la communication entre les deux entités peut affecter les performances de mémorisation. L'expérience utilisée pour vérifier cette hypothèse peut vouloir montrer que le fait d'intercaler une tâche de répétition de chiffres entre la phase d'encodage d'une liste de mots et la phase de rappel fait baisser sensiblement la performance du participant sur les derniers items sans modifier les résultats sur le reste de la liste.

Les modèles mathématiques, sous la forme d'équations, fournissent une formalisation des modèles théoriques auxquels ils sont associés. De plus, ils peuvent mener à une implémentation et, tout comme les modèles computationnels, ils permettent de tester et de valider les différentes hypothèses sous-jacentes et également les limites d'utilisation. Cependant leur pouvoir d'expression reste limité : ils caractérisent finement les relations entre les entrées et les sorties mais ne peuvent décrire aisément les mécanismes sous-jacents. Ils ne peuvent pas par exemple modéliser le stockage ou le transfert d'informations. Ils sont donc en quelque sorte à la frontière entre les modèles théoriques et computationnels.

Dans cette thèse nous nous intéressons à cette dernière catégorie de modèles. Les modèles computationnels, car ce type de modèle cognitif a la capacité de simuler, autant que faire se peut, le comportement humain, et de produire des données comportementales comparables à celles des humains. De plus, l'implémentation mise en place dans cette modélisation implique une réflexion de fond sur les théories inhérentes à la modélisation des processus, afin d'en extraire une spécification algorithmique. Enfin, une fois le modèle conçu et validé comme proche du comportement humain, son exécution, en jouant par exemple sur les paramètres, peut aussi accroître nos connaissances des processus cognitifs simulés.

Ainsi un des objectifs de la modélisation, telle que nous l'entendons ici, est d'être le plus simple possible et de limiter le nombre de paramètres utilisés. Le but est d'atteindre un équilibre entre cette simplicité et la reproduction des données. Un grand nombre de paramètres peuvent certes permettre de reproduire les données empiriques lors d'une tâche particulière, mais le modèle risque d'être trop ad-hoc et de ne pas être robuste au moindre changement dans la tâche ou les stimuli. En étant trop proche des données un modèle perd de son pouvoir prédictif (Rissanen, 1978). À l'inverse un modèle trop simple n'est pas assez précis pour simuler les comportements et ne permet ni de comprendre ces comportements ni d'être utilisé en terme de prédiction. Dans notre démarche de modélisation, nous essayerons donc de produire un modèle capable de se comparer aux performances humaines tout en

limitant le nombre de paramètres.

Les modèles cognitifs computationnels peuvent être intégrés dans une architecture cognitive. Dans ce cas, la cognition humaine est représentée sous forme d'une structure comportant différents modules liés entre eux. Ces modules représentent les processus moteurs (mouvements des yeux, clic souris) et perceptifs (vision, audition), mais aussi des processus mnésiques (mémoire déclarative et procédurale). Ces architectures, comme ACT-R (Anderson et al., 2004), EPIC (Kieras & Meyer, 1997) ou SOAR (Newell, 1990), donnent un cadre de base pour l'étude des processus cognitifs sur des tâches particulières. L'avantage de ces architectures est de posséder des modules validés par des études préalables et donc même si le nombre de paramètres est important ceux-ci ont été fixés suite à différentes expérimentations sur différentes tâches.

Afin d'avoir une plus grande liberté dans les choix de modélisation, au niveau de la séparation entre les processus, et des liens qu'ils entretiennent, d'autres modèles existent, ne se basant pas sur une telle architecture. Nous nous placerons dans cette catégorie de modèles et nous nous concentrerons sur un aspect de la cognition. Le système cognitif humain étant en effet d'une telle complexité qu'une méthode consiste à se focaliser sur un point précis afin d'essayer de le comprendre. L'inconvénient de ces modèles est qu'étant moins testés et validés par ailleurs, ils sont plus sensibles au risque de multiplication des paramètres évoqué précédemment.

Nous nous intéressons ici à la recherche d'information qui est une activité cognitive se retrouvant très largement dans notre vie quotidienne, et plus particulièrement lorsque cette recherche se situe dans des documents structurés, composés d'éléments textuels et graphiques. Ces documents, intégrant des textes et des images présentés simultanément, peuvent être des pages de journaux, de magazines, des publicités et plus particulièrement des pages web. Une des tâches les plus fréquentes sur le web est en effet de rechercher des informations, par exemple chercher la meilleure solution parmi toutes celles proposées par un moteur de recherche, ou plus généralement un article sur un sujet précis, ou simplement un mot dans une page. Cette quête d'information peut être plus ou moins précise. Dans le cas d'une recherche dont nous connaissons exactement le ou les mots la plupart des applications comportent aujourd'hui un module de recherche permettant de signaler ces mots à l'utilisateur (en utilisant le raccourci CTRL-F, voir figure 2 page suivante), et il s'agit plutôt d'une recherche visuelle. Par exemple si nous voulons trouver un nom dans un annuaire en ligne, il nous suffit de taper ce nom pour le mettre en évidence dans la page (en général en surligné et centré). Dans le cas de recherche moins précise, par exemple rechercher un article sur le *voilier* mais sans impliquer forcément que l'article comporte ce mot, une recherche plus approfondie de l'information dans la page est nécessaire. C'est cette recherche que nous nous efforcerons de simuler, et plus particulièrement les mouvements oculaires produits lors de cette activité cognitive à un niveau de précision suffisamment fin ; au niveau des fixations et des saccades oculaires. Nous nous plaçons donc ici dans le cadre de la vision active (*active vision*) qui stipule



FIGURE 2 – Zone de recherche dans une application.

que les mouvements oculaires sont un aspect crucial de notre interaction visuelle avec le monde, et donc décisifs dans le cadre de la recherche d'information (Findlay & Gilchrist, 2003). La technique d'enregistrement des mouvements oculaires nous permettra une investigation des processus mis en jeu dans cette fonction cognitive complexe.

Lors de cette tâche cognitive deux types d'attention sont déployés, l'attention *overt* et l'attention *covert*. La première est observable et les mouvements oculaires utilisés comme reflet de la cognition sont ici associés à l'attention *overt*. L'attention *covert* au contraire n'est pas observable ni quantifiable objectivement même si son existence est indéniable (Findlay & Gilchrist, 1998). Des éléments de la page non fixés, c'est-à-dire ne recevant aucune fixation oculaire, peuvent donc être traités par un déplacement de l'attention *covert*, mais nous ne prendrons pas ce déploiement de l'attention en compte étant donné le manque de moyens objectifs d'évaluation de ce type de prise d'information.

La méthodologie que nous utilisons dans cette thèse autour de la conception d'un modèle computationnel est la suivante. La première étape consiste à identifier les fonctions cognitives jouant un rôle dans le comportement que nous cherchons à modéliser. Une fois ces fonctions mises en évidence, un recueil des informations autour de ces fonctions est nécessaire. Ces informations viennent de données physiologiques, de modèles théoriques ou d'autres modèles computationnels. La conception du modèle, basée sur ces informations, intervient de manière à la fois simple et cognitivement plausible. Si besoin, les paramètres du modèle sont fixés. Le modèle est alors prêt à être exécuté, mais nous ne pouvons espérer modéliser le comportement humain sur une tâche qu'à partir du moment où il y a une cohérence dans les données comportementales humaines. Cette cohérence peut se retrouver à plusieurs niveaux de granularité différents. Par exemple des trajets oculomoteurs pris au niveau des fixations peuvent être très hétéroclites, mais au niveau des transitions entre éléments dans la page une cohérence peut émerger. Si nous pouvons caractériser les performances humaines, avec des variables globales, alors à ce niveau d'abstraction nous pourrions comparer les modèles et les humains. Nous voyons donc le besoin d'analyser les comportements humains préalablement à la comparaison avec les simulations afin de cibler les variables stables chez les humains. Une fois ces variables identifiées, nous pouvons passer à la comparaison entre les données comportementales et le modèle. Une phase d'expérimentations dans laquelle les participants sont soumis au même matériel et aux mêmes tâches que le modèle permet de récupérer les données empiriques à comparer. La comparaison entre les données expérimentales et simulées, basée sur les variables de haut niveau identifiées préalablement, est donc la dernière étape à mettre en place.

Modéliser pour prédire

Rendre compte correctement des mouvements oculaires lors de la recherche d'information pourrait permettre non seulement une meilleure connaissance des processus sous-jacents à la recherche d'information mais aussi une meilleure connaissance de la manière dont nous utilisons les interfaces, comme l'expliquent Halverson et Hornof (2007).

Un tel outil peut donc faciliter la caractérisation de l'utilisabilité des interfaces. À côté de méthodes qualitatives (entretiens, questionnaires ou critères théoriques), les méthodes d'évaluation de l'utilisabilité d'une interface fondées sur un modèle cognitif de l'utilisateur permettent une meilleure prédiction des performances (taux d'erreurs, séquences d'actions, durées de traitement) (Dix, Finlay, Abowd, & Beale, 1998 ; Baccino, Bellino, & Colombi, 2005), mais également des mesures physiologiques (Salvucci, 2001). Ces types de modèles prennent en compte les caractéristiques cognitives des futurs utilisateurs pour véritablement simuler leurs comportements et détecter le plus tôt possible dans le processus de conception de l'interface des problèmes d'utilisabilité (West & Emond, 2001), tout en rendant explicite la nature de ces problèmes (Byrne, 2005). Cette méthodologie possède une longue histoire, depuis les anciens modèles GOMS (Card, Moran, & Newell, 1983) ou CCT (Kieras & Polson, 1985) jusqu'à des modèles plus récents, par exemple CoLiDeS (Kitajima, Blackmon, & Polson, 2000) ou SNIF-ACT (Pirolli & Fu, 2003) basé sur l'architecture ACT-R. Elle requiert une expertise très fine dans le domaine de la cognition, afin de prédire au mieux les comportements humains. Le but n'est pas de concevoir un système informatique qui traiterait un document plus vite et mieux que ne le feraient des humains (c'est le cas du résumé automatique de documents par exemple), mais bien de coller au plus près des comportements humains, dans un objectif, entre autres de caractérisation de l'utilisabilité de l'interface. Des applications de cette approche existent dans le domaine industriel essentiellement pour la prédiction des durées de traitement et des probabilités d'erreur (Ritter & Young, 2001). Des modélisations plus ambitieuses en sont encore au stade du prototype de recherche, comme par exemple CogTool (John & Salvucci, 2005) et CORE/X-PRT (Tollinger et al., 2005), et les perspectives pratiques des simulations d'utilisateurs sont réelles.

Processus modélisés

Le but de cette thèse est d'esquisser un modèle cognitif intégrant les processus mis en jeu lors d'une tâche de recherche d'information. Les processus que nous avons identifiés sont visuels, sémantiques et mnésiques.

La recherche d'information suppose le traitement de deux principales sources d'information : les informations visuelles et les informations sémantiques. Les premières font référence à des caractéristiques visuelles, traitées par des processus bas niveau, ascendants, tandis que les secondes font référence à la signification du contenu, et

font appel à des processus de haut niveau, descendants. Les éléments attirant l'attention peuvent donc être saillants sur deux plans : leur saillance peut être perceptive (caractéristiques physiques des éléments, liées à un processus ascendant) ou cognitive (dépendant des processus cognitifs descendants) (Landragin, 2004). En effet, un document structuré (comme une page Web) peut être vu comme la réunion de deux types d'entités : des unités textuelles multiples (à la différence d'une seule page de texte) et des caractéristiques visuelles diverses (couleurs, images, mises en forme, etc.). Ces deux caractéristiques correspondent à deux champs de recherche assez indépendants que nous voudrions réunir dans une même modélisation : la modélisation de la compréhension de textes, depuis les modèles de lecture (Kintsch, 1998) jusqu'aux modèles de traitement sémantique (Lemaire, Denhière, Bellissens, & Jhean-Larose, 2006 ; Landauer, McNamara, Dennis, & Kintsch, 2007), qui sont habituellement appliqués à des textes uniques et la modélisation des processus visuels (Itti & Koch, 2000), traditionnellement appliqués à des scènes visuelles sans contenu textuel. Les interactions entre les processus visuels et les processus sémantiques constituent donc la spécificité de notre approche.

La plupart des modèles computationnels de sélection de l'attention visuelle sont basés sur la théorie d'intégration des traits (Treisman & Gelade, 1980). Le plus connu est celui proposé par Itti et Koch (2000), basé sur une décomposition des caractéristiques visuelles des stimuli (couleur, fréquence, orientation). Pour les premières fixations et dans le cas d'exploration libre, ces modèles concordent avec les données expérimentales. Cependant au cours du temps et lors d'une tâche cognitive, l'influence des processus cognitifs descendants devient prépondérante sur les processus visuels bas-niveau ascendants (Henderson, 2007). En raison de cette complexité, peu de modèles sont proposés intégrant cette modulation des processus ascendants et descendants. Notons néanmoins l'architecture générique proposée par Navalpakkam et Itti (2005), que nous détaillerons ultérieurement.

Au niveau sémantique peu de modèles computationnels de recherche d'information intègrent des informations visuelles. Avec l'Analyse de la Sémantique Latente (LSA) (Landauer et al., 1998), une représentation des connaissances peut être implémentée et estimée à partir de l'analyse automatique de grands corpus de textes. Par exemple, dans le modèle CoLiDeS les auteurs utilisent LSA dans ce cadre et décrivent la navigation dans des pages web à partir de la similarité entre les objets (boutons, liens) et le but de l'utilisateur (Kitajima et al., 2000). Mais ce modèle n'intègre pas de prise en compte d'informations visuelles.

Un tel modèle de recherche d'information serait incomplet sans un mécanisme de mémoire de travail. Celle-ci permet de garder une trace du trajet effectué et empêche l'attention de retourner sur des localisations traitées auparavant. Différents modèles seront utilisés dans cette thèse, le mécanisme d'inhibition de retour (IOR) (Klein & MacInnes, 1999) et le *Variable Memory Model* (Arani, Karwan, & Drury, 1984 ; Horowitz, 2006).

Notre problématique réside donc dans cette intégration à la fois au niveau théo-

rique et au niveau pratique. Au niveau théorique, ces modèles ne partagent pas les mêmes origines ni les mêmes objectifs et nous devons donc parcourir une vaste littérature afin d'intégrer les connaissances nécessaires à la réalisation d'un modèle regroupant les différentes parcelles de ce champ interdisciplinaire. Au niveau pratique nous étudierons les différents modèles théoriques et computationnels qui traitent de recherche d'information, de vision, de sémantique et de mémoire, et notre implémentation les intégrera dans une seule et même structure.

Contexte de travail

Cette thèse s'est déroulée dans le cadre d'une collaboration avec Thierry Baccino, professeur de psychologie cognitive, au sein du projet ANR Eye-LSA. Les objectifs de ce projet étaient de concevoir un modèle cognitif computationnel simulant la recherche d'information sur le web (en combinant les traitements visuels et sémantiques) à partir de tâches réalistes et de buts mal définis, et plus précisément le comportement visuel d'un utilisateur engagé dans une recherche d'information avec une validité psychologique satisfaisante.

Une collaboration avec le laboratoire *Visual Cognition Laboratory* de John Henderson à Edinburgh, lors d'un séjour doctoral de six mois, a aussi été établie.

Contenu de la thèse

Cette thèse s'articule en six chapitres. Le premier chapitre sur la recherche d'information présente des modèles de recherche d'information et de lecture proches de notre approche. Le second chapitre décrit notre modèle. Ce modèle est un modèle spatio-temporel basé sur l'intégration dynamique des informations visuelles et sémantiques associées à un mécanisme mnésique. Le chapitre trois a pour ambition de fournir des éléments de bases ré-exploitable dans la production d'un modèle plus complexe applicable à des pages plus riches. Les situations très contrôlées qui y sont décrites nous ont permis de vérifier les paramètres de notre modèle. Dans les chapitres quatre, cinq et six, le modèle est appliqué et adapté à des pages plus complexes intégrant plusieurs niveaux de navigation dans la page. Nous finirons ensuite par une discussion générale et les perspectives de recherches futures suite à ce travail.

1

Modélisation cognitive de la recherche d'information

Dans cette thèse, nous nous placerons dans le cadre de la modélisation cognitive des processus mis en œuvre par un humain lors d'une tâche de recherche d'information. Cette vue de la recherche d'information en tant que mécanisme cognitif complexe se rapproche de celle d'*information seeking* proposée par Marchionini (1997) et celle d'*information foraging* développée par Pirolli et Fu (2003). Aussi le sens que nous donnons à la recherche d'information diverge de celui couramment utilisé en informatique : nous chercherons à développer un modèle capable de retrouver une information d'une manière similaire à celle des humains, c'est-à-dire le plus cognitivement plausible, et non pas le plus rapidement possible.

Quels sont les mécanismes qui guident l'attention visuelle lors de l'exploration de notre environnement ? Une distinction classique est faite entre les mécanismes dit ascendants (*bottom-up*) et les mécanismes descendants (*top-down*). Les premiers relèvent d'un traitement des informations exogènes, caractéristiques visuelles de l'environnement, et les seconds d'un traitement des informations endogènes, c'est-à-dire les connaissances, les objectifs et les attentes de l'observateur. Ces deux types de traitements sont fortement liés, et nous ne saurions imaginer un système cognitif complet sans l'un des deux ou même sans interaction. Prenons comme exemple pour illustrer cette complémentarité l'interprétation des ombrages.

Ramachandran (1988) s'est intéressé à la façon dont nous interprétons des stimuli ombrés. Dans la figure 1.1 page suivante, deux interprétations en trois dimensions sont possibles pour chaque cercle. Chaque interprétation dépend de l'hypothèse sur

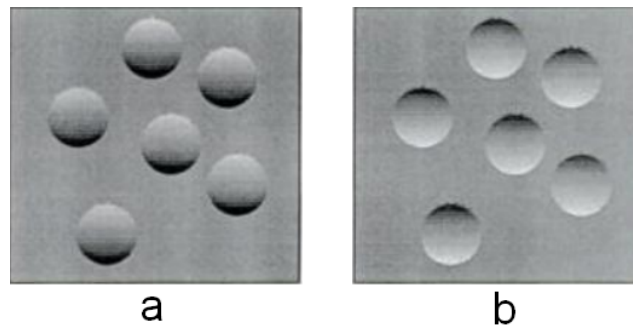


FIGURE 1.1 – (a) Stimuli interprétés de manière générale comme des bosses. (b) Stimuli interprétés de manière générale comme des creux. (Ramachandran, 1988)

l'origine de la lumière choisie (le cercle peut être vu comme un creux ou une bosse). Cette simple tâche permet de mettre en évidence la contribution des processus descendants car la plupart des participants de l'expérience font l'hypothèse que la lumière vient d'en haut. Ce choix résulte de connaissances sur le monde dans lequel nous vivons où la lumière (le soleil ou les lumières artificielles) vient généralement du haut.

Le guidage de l'attention se ferait donc à la fois au niveau des propriétés visuelles (connaissances des caractéristiques visuelles) et au niveau conceptuel (connaissances des caractéristiques sémantiques). Ce constat était déjà d'actualité il y a plusieurs décennies dans les travaux de Yarbus (1967), celui-ci ayant donné comme consigne à des observateurs d'estimer le statut social de la famille ou de donner l'âge des personnages d'une scène. Dans un cas, les participants regardent en priorité les vêtements, tandis que dans le second cas ce sont surtout les visages qui sont fixés. Nous voyons avec cet exemple l'influence qu'a la tâche dans le guidage de l'attention.

Pour cela, nous commencerons donc par décrire les modèles de recherche visuelle ascendants, puis ceux intégrant des processus descendants. Nous verrons ensuite différents modèles de lecture et nous finirons par deux modèles cognitifs de recherche d'information : SNIF-ACT et CoLiDeS.

1.1 Guidage ascendant de l'attention

Un mécanisme ascendant du guidage de l'attention apparaît nécessaire à l'être humain pour réagir et adapter son comportement face à son environnement. Nous imaginons aisément l'homme des cavernes face à ses prédateurs devant être alerté au moindre mouvement, ou l'homme moderne, face aux phares d'un véhicule arrivant dans la nuit. Dans ces situations, il apparaît en effet préférable d'avoir automatiquement son attention dirigée vers ces éléments, qui sont dits saillants. Mais leur saillance n'est pas une propriété en soit. Par exemple, un trait vertical au milieu de traits plus ou moins horizontaux sera saillant, mais au milieu d'autres éléments plus ou moins verticaux son orientation ne constituera plus un trait saillant dans le

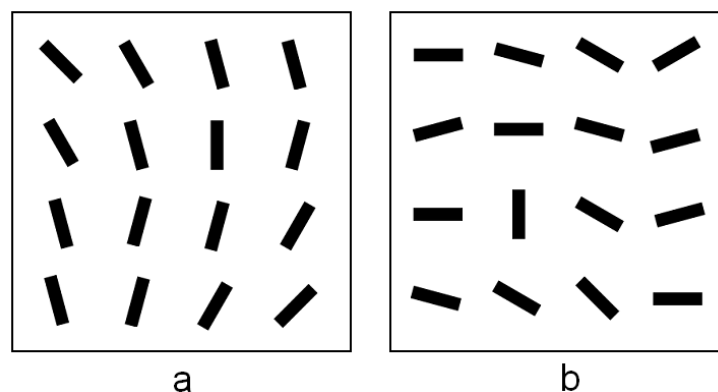


FIGURE 1.2 – (a) L'élément vertical n'est pas saillant visuellement car son entourage a des orientations proches. (b) L'élément vertical est saillant car son orientation le différencie de son entourage.

champ visuel. La saillance d'un élément dépend donc du contexte dans lequel il se situe, des éléments environnants (voir figure 1.2).

Ce guidage ascendant de l'attention a encouragé les recherches et la modélisation. Ainsi des modèles de saillance visuelle basés sur les caractéristiques visuelles des stimuli ont vu le jour.

La théorie à la base de tous les modèles générant des cartes de saillance est la théorie d'intégration des traits (*feature-integration theory of attention*) de Treisman et Gelade (1980).

1.1.1 Théorie d'intégration des traits

Dans leurs travaux, les auteurs ont mis en évidence deux étapes de traitement visuel de l'image source lors de recherche visuelle : une étape pré-attentive et une étape attentive. La première étape, qui s'effectue en parallèle, permet l'extraction des caractéristiques de bas niveau de la scène (orientation, couleur, contraste...). Il résulte de cette étape une carte de saillance de la scène, où les éléments saillants entraînent un effet *pop-out*. Cette représentation de l'image dans ces dimensions permet par exemple d'expliquer la ségrégation de texture.

La seconde étape est en série, et permet l'intégration des traits pour former des objets à partir des caractéristiques de bas niveau. L'attention se déplace afin de se focaliser sur chaque objet de la scène et ce focus attentionnel agit comme une « glu » qui intègre les caractéristiques, initialement séparées, en un objet unifié. Dans cette seconde phase, des informations descendantes, issues du contexte et des expériences passées, entrent en jeu pour former une perception correcte des éléments.

À partir de cette théorie une première implémentation a été créée par Koch et Ullman (1985). Ce modèle computationnel a été repris, amélioré et popularisé par Itti et Koch (2000, 2001). Le principe de ce modèle est de générer des cartes topographiques de saillance visuelle à partir d'images (ou de vidéos (Itti, 2004)).

1.1.2 Modèle de cartes de saillance de Itti

Dans ce modèle, les cartes de saillance sont construites à partir de l'extraction dans différentes dimensions des traits visuels des images originales. Pour chaque dimension une carte spatiale est générée et une intégration de ces dimensions se fait ensuite pour générer la carte de saillance finale. Un mécanisme local de recherche de maxima successifs (*Winner Take All*) permet de sélectionner la région présentant une saillance maximale et devenant une cible privilégiée pour un déplacement, par une saccade, de l'attention visuelle en ce site. À partir de cette carte, il est donc possible de générer un parcours plausible de scrutation de l'image où la séquence des fixations oculaires suit la séquence des régions ainsi sélectionnées. La première région sélectionnée est fixée, elle est ensuite inhibée par un mécanisme d'inhibition de retour (IOR), une nouvelle région est sélectionnée, et ainsi de suite.

Les trois principales dimensions prises en compte dans ce modèle sont la couleur, l'intensité, et l'orientation. D'autres dimensions peuvent être prises en compte, comme le mouvement, les contours, les bords ou les ombres, voir figure 1.3 page ci-contre. Sur cette figure nous pouvons voir à partir de l'image d'entrée (en haut à gauche) l'extraction de toutes les caractéristiques de bas niveau : les différents canaux de couleur, les différentes orientations, etc. À partir de chacune de ces cartes des pyramides gaussiennes à plusieurs échelles sont construites avec un filtrage passe-bas et un sous-échantillonnage qui permettent de supprimer les détails et de ne garder que les régions les plus saillantes dans chaque dimension. La dernière étape est une normalisation de toutes les images qui sont ensuite intégrées pour former la carte de saillance finale (en bas à gauche). La région la plus saillante est ensuite sélectionnée (*winner take all*). (voir (Itti, Koch, & Niebur, 1998) pour les détails mathématiques).

Ce modèle a été utilisé dans des applications telles que la recherche d'information lors d'une tâche de conduite (repérage de panneaux et obstacles), la détection de cibles militaires et l'évaluation de couvertures de magazines de mode. Les prédictions de ce modèle correspondent bien aux performances des humains dans les tâches simples de détection de cible, lorsqu'il est appliqué à des images à faible contenu sémantique.

Lorsque l'environnement devient plus riche, et que nous nous éloignons des conditions expérimentales très épurées, ces effets de saillance purement ascendants diminuent et les modèles de saillance visuelle simplement basés sur les caractéristiques visuelles de la scène ne permettent plus d'expliquer le comportement humain. Ainsi Henderson, Brockmole, Castelhana, et Mack (2007) ont montré que les mouvements oculaires des participants effectuant une tâche de recherche visuelle sur des photographies en couleur du monde réel, ne correspondaient pas aux prédictions faites par le modèle d'Itti et Koch. Comment expliquer par exemple sans mécanisme descendant qu'un cycliste sur un toit attirera plus l'attention que sur une piste cyclable ? Dans le cas de cet exemple c'est le caractère impromptu de sa présence (niveau sémantique) qui le rend saillant et non ses caractéristiques visuelles.

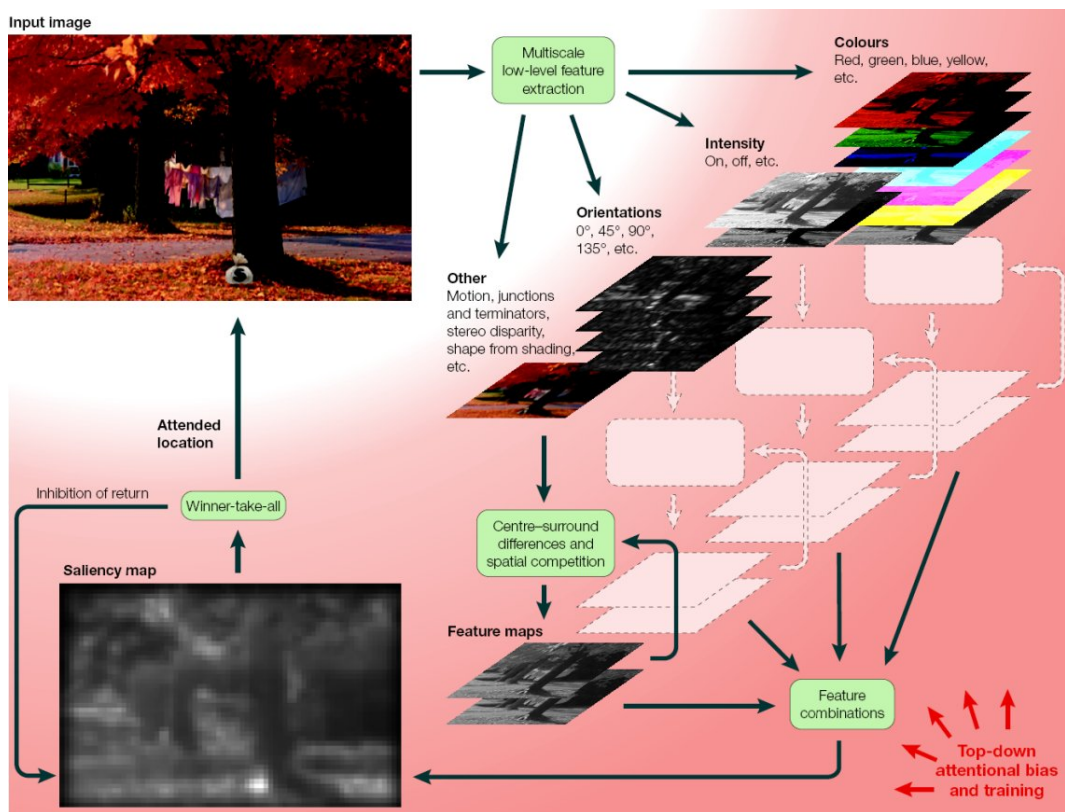


FIGURE 1.3 – Modèle de génération de carte de saillance (Itti & Koch, 2001).

1.2 Modèles de recherche visuelle

Comme nous venons de le voir, les mécanismes ascendants de guidage de l'attention (guidés par les données) ne permettent pas de rendre compte du comportement humain dans des tâches complexes de recherche visuelle. Les connaissances acquises, les attentes et les buts de l'individu sont les principales sources d'information qui vont guider de manière descendante le déplacement de l'attention visuelle. Par exemple, lors de la recherche d'un objet connu, les traits visuels de cet objet vont être pré-activés, afin de faciliter la recherche (Duncan & Humphreys, 1989). Les connaissances que nous avons de cet objet vont aussi permettre de limiter la recherche à certaines localisations. Par exemple, si nous cherchons une voiture noire dans un parking, même si une voiture rouge pourrait paraître saillante elle n'attirera pas notre attention car elle ne partage pas ce trait visuel avec la cible, qui a été induit par une consigne ou une intention.

Cette influence des traitements descendants fut l'objet de modélisations théoriques, entre autres par Desimone et Duncan (1995) avec leur modèle de compétition biaisée de la sélection visuelle. D'après ces auteurs, des signaux descendants semblent biaiser la sélection en faveur de l'objet pour lequel les traits ont été pré-activés et qui gagne donc la compétition sélective entre les objets dans le champ visuel. Ce modèle de compétition biaisée a été augmenté par Moores, Laiti, et Chelazzi (2003)

qui lui ont ajouté une prise en compte d'information sémantique. Dans leurs expériences, les objets sémantiquement proches de la cible étaient plus susceptibles d'être sélectionnés que les autres.

Les modèles actuels de guidage de l'attention prennent donc en compte à la fois les informations exogènes et endogènes de l'environnement pour décrire le comportement humain.

Dans cette section, nous nous attacherons à décrire des modèles computationnels de recherche visuelle s'appliquant à des environnements simples et complexes. Le premier est le modèle *Guided search 4.0* développé par Wolfe (2007), basé sur les caractéristiques visuelles des éléments présents, qui intègre des processus ascendants et descendants, mais ne s'applique qu'à des environnements simples. Le second est le modèle de Pomplun (2007), *Area activation*, qui peut aussi être appliqué à des scènes naturelles. Nous verrons ensuite le modèle de Navalpakkam et Itti (2005), appliqué à des scènes complexes et basé sur des cartes de saillance visuelle.

1.2.1 Modèle *Guided Search*

L'objectif du modèle *Guided Search 4.0* (Wolfe, 2007) est de rendre compte de la capacité humaine à identifier un objet parmi un ensemble de distracteurs. Afin de décrire comment l'attention est dirigée vers le bon élément, ce modèle combine des mécanismes ascendants de saillance visuelle (basée sur des caractéristiques telles l'orientation et les couleurs) et des mécanismes descendants prenant en compte les informations extérieures liées à l'élément recherché.

Sur la figure 1.4 page suivante, si la tâche est de rechercher les lignes horizontales, les mécanismes ascendants vont entrer en conflit avec les mécanismes descendants par rapport à la ligne verticale qui se trouve au milieu de la figure, de part sa saillance visuelle (orientation différente de celle de ses voisins), mais qui n'est pas pertinente pour la recherche. Cela peut être modélisé par une pondération entre les mécanismes ascendants et descendants. En mettant par exemple le poids des mécanismes ascendants à 0, l'effet sera d'ignorer totalement les éléments saillants (dont les non pertinents). Au contraire en ne permettant pas ce poids nul, ce qu'ont choisi les auteurs, le modèle offre la possibilité de déplacer son attention sur des éléments non pertinents pour la recherche.

Dans ce modèle, une carte est générée dans laquelle l'activation des éléments augmente si la dissemblance entre l'élément et ses voisins augmente (mécanisme ascendant) et si sa ressemblance avec la cible augmente (mécanisme descendant). À partir de ce double mécanisme d'activation, il est alors possible de générer un trajet de déplacement de l'attention. Elle est tout d'abord attirée par l'élément ayant la plus forte activation, puis par le second, etc. Ce modèle génère néanmoins des trajets différents de ceux des humains car les saccades peuvent être très grandes et désordonnées contrairement aux observations empiriques. De plus la tâche simulée par cette modélisation est de retrouver une cible connue, et il n'y a donc pas d'effet sémantique.

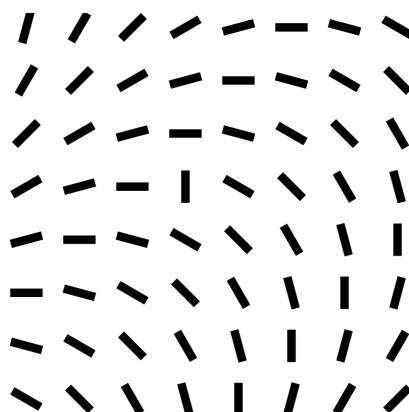


FIGURE 1.4 – Exemple d’application du modèle *Guided Search* (Wolfe, 2007)

1.2.2 Modèle *Area Activation*

Le modèle développé par Pomplun (2007) répond à une des critiques que nous avons pu faire à celui de Wolfe dans la mesure où il tient compte d’un coût pour les longues saccades et tente donc de réduire au minimum la longueur du trajet. À partir d’une fixation la suivante est calculée afin d’être proche de la fixation courante.

Ce modèle peut être appliqué à des scènes naturelles, sur le même principe que celui de Wolfe. À partir d’une cible prise dans l’image d’origine (un fragment d’image), un double mécanisme va moduler l’activation des zones selon la distance avec les éléments locaux (mécanisme ascendant) et la proximité avec la cible (mécanisme descendant). À ce mécanisme se greffe cette notion de minimisation de la longueur du trajet.

La figure 1.5 page suivante présente l’image avec comme cible de la recherche la barre verticale grise, la carte d’activation simulée à partir des informations ascendantes et descendantes et le trajet simulé par le modèle lors de cette recherche visuelle.

Ce modèle ne prend en compte que les caractéristiques visuelles des éléments sans permettre d’ajouter des informations sémantiques à la recherche.

1.2.3 Modèle de Navalpakkam et Itti

Un des modèles qui tient compte de la double prise d’informations : visuelles et sémantiques, est celui de Navalpakkam et Itti (2005) que nous allons détailler. D’autres modèles vont aussi dans cette direction, comme celui de Tatler, Baddeley, et Gilchrist (2005), ou celui de Torralba, Oliva, Castelhana, et Henderson (2006). Le point commun de tous ces modèles est de partir de cartes de saillance auxquelles des contraintes descendantes sont appliquées.

L’objectif de ce modèle est de simuler les déplacements de l’attention dans une scène lors d’une recherche de cible, en prenant en compte à la fois des informations endogènes et exogènes. Au niveau des informations endogènes, le modèle se base sur

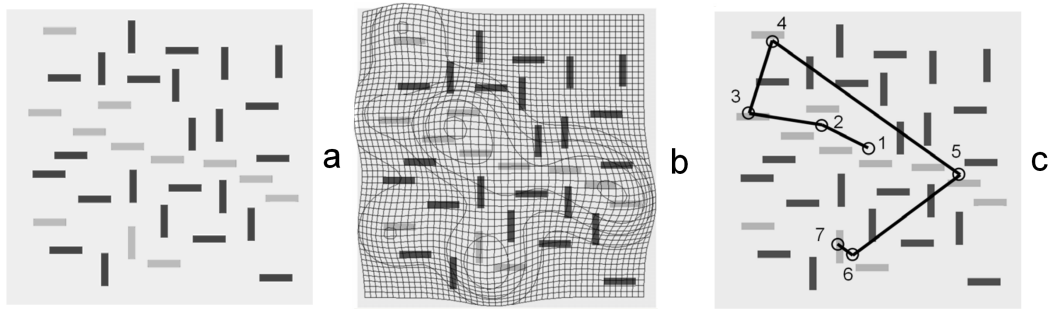


FIGURE 1.5 – Exemple d'application du modèle *Area Activation* (Pomplun, 2007). (a) image, cible : barre verticale grise (b) carte d'activation simulée à partir des informations ascendantes et descendantes (c) trajet simulé par le modèle.

celui de Itti et Koch (2000) : si la cible est rouge, la carte de saillance va être modulée de façon à privilégier les pixels rouges. Au niveau des informations exogènes, si par exemple l'objet recherché est une personne dans une scène de plage (avec plage, mer et ciel) la partie ciel sera pénalisée, car des informations en mémoire à long terme permettront au modèle de savoir qu'il est rare de voir des personnes dans le ciel.

L'exécution du modèle de Navalpakkam et Itti se déroule en quatre phases :

Phase 1 : Les yeux fermés. Dans cette première phase il s'agit de la modélisation de ce qui est su indépendamment des stimuli visuels. Lors de cette phase, la mémoire de travail de l'utilisateur est initialisée à partir de la définition de la tâche. La mémoire de travail a accès à la mémoire à long terme pour faire cette initialisation. Si par exemple la tâche est de trouver ce qu'un personnage est en train de manger, des connaissances sur le fait que se nourrir est lié à une action entre la bouche et la main aidera à orienter l'attention vers la bouche ou la main pour déterminer l'aliment. L'utilisation de ces connaissances préalables par la mémoire de travail permet de déterminer les éléments pertinents pour la recherche et comment ils sont liés les uns aux autres. Les connaissances dans la mémoire à long terme sont codés manuellement sous forme d'ontologie. De cette phase est générée une carte de la tâche.

Phase 2 : Computation. Dans la deuxième phase, « les yeux sont ouverts » et le système visuel du modèle reçoit la scène en entrée. La carte de saillance visuelle ascendante modulée par les caractéristiques visuelles de la cible est calculée. Cette carte de saillance modulée est ensuite combinée à celle de la tâche pour donner une carte de guidage de l'attention. Un mécanisme de sélection de la meilleure région (*Winner Take All*) permet de sélectionner la région où l'attention va se focaliser.

Phase 3 : Vérification. Dans la troisième phase, le modèle de reconnaissance d'objets détermine l'identité de l'entité fixée, et estime si la tâche est effectuée ou non.

Phase 4 : Mise à jour. Cette phase consiste à inhiber dans la mémoire de travail

de la localisation fixée si celle-ci n'avait pas répondu à la tâche, puis mettre à jour des cartes.

Les étapes 2, 3 et 4 se répètent jusqu'à la réalisation de la tâche.

Ce modèle est donc très intéressant car il montre bien l'avantage d'ajouter un traitement sémantique lors d'une recherche d'information visuelle. La principale limite de ce modèle est la mémoire sémantique à long terme (sous forme d'ontologie) qui est codée manuellement. Les images doivent par ailleurs subir une phase de segmentation manuelle des objets ainsi qu'une labélisation de ces objets pour pouvoir les inclure dans l'ontologie. L'ontologie ainsi créée n'est donc pas exhaustive et demande pour chaque nouveau stimuli ou cible d'être remise à jour. De plus les connaissances ainsi représentées dépendent de la personne qui les a codées et n'ont pas de validation en tant que modèle des connaissances sémantiques en mémoire à long terme par des utilisateurs.

1.3 Modèles de recherche dans des documents textuels

Les différents modèles que nous avons vu jusqu'alors s'attachaient à modéliser la recherche d'information et le guidage de l'attention dans des environnements artificielles, ou des scènes naturelles. Or nous cherchons à modéliser la recherche d'information dans des documents structurés comme des pages web contenant des informations sous forme textuelle en plus d'informations visuelles. Ce type d'information entraîne de nouveaux processus : lecture, compréhension de texte, stratégies de navigation entre régions informatives, etc. Nous allons donc étudier quels sont les modèles associés à ces différentes fonctions cognitives.

1.3.1 Modèles de lecture

Lors du traitement de documents textuels, la question de la lecture et de la compréhension de texte apparaît essentielle. Nous allons donc décrire trois des principaux modèles de lecture de la littérature. Ces modèles s'attachent à simuler les trajets oculomoteurs lors d'une tâche de lecture linéaire de phrases ou de paragraphes.

Nous distinguons classiquement deux types d'approche dans la modélisation des mouvements oculaires en lecture : les modèles cognitifs, qui partent du principe que les mouvements oculaires dépendent en premier lieu des propriétés langagières des mots, et les modèles de contrôle oculomoteur primaire (*primary oculomotor control*, POC) qui eux privilégient les informations bas niveau des mots, sans prendre en compte la notion d'attention. Les trois modèles que nous décrirons ici sont des modèles cognitifs. Dans ces modèles cognitifs nous pouvons distinguer deux classes : les modèles assumant une distribution de l'attention comme un gradient durant la lecture (GAG, *guidance by attentional gradient*) et les modèles assumant une allocation sérielle de cette attention d'un mot à l'autre durant la lecture (SAS, *sequential attention shift*).

***EZ-Reader* (SAS)**

EZ-Reader est un modèle qui décrit la façon dont la perception, la cognition, et les propriétés oculomotrices déterminent *où* et *quand* les yeux se déplacent pendant la lecture (Reichle, Pollatsek, Fisher, & Rayner, 1998 ; Reichle, Rayner, & Pollatsek, 2006).

Les principales hypothèses du modèle sont que les mots sont identifiés un par un au cours de la lecture, et que le processus d'identification des mots est le « moteur » qui anime les yeux au cours de la lecture. Ce modèle a permis de répondre à plusieurs questions portant sur la lecture, comme par exemple la manière dont la syntaxe et/ou la sémantique influencent le traitement des mots. Nous noterons par exemple que les mots-outils (non porteurs de sens, comme les déterminants) sont moins fixés que les mots plus informatifs, que les mots plus fréquents sont fixés moins longtemps, ou encore que les mots ayant plusieurs sens sont fixés plus longtemps.

EZ-Reader comporte trois étapes pour l'identification des mots :

- L'analyse visuelle précoce (V), pré-attentive ;
- puis deux étapes lexicales (étapes attentives) :
- L1 : vérification de la familiarité (*familiarity check*) qui permet la programmation d'une saccade vers le mot n+1 ;
 - L2 : achèvement de l'accès au lexique (*completion of lexical access*) suivi d'un déplacement de l'attention vers le mot n+1 (déplacement du spot attentionnel).

Le dépassement d'un certain seuil par le niveau de familiarité du stimulus (par rapport aux représentations en mémoire) serait le signal qui indiquerait au système de programmation des saccades que l'accès lexical est imminent et qu'une saccade peut être planifiée en direction du mot suivant. Ce modèle prend également en compte l'excentricité des mots parafovéaux qui influence l'efficacité du traitement, en préparant la saccade suivante à partir des informations bas-niveau sur les frontières des mots. Il intègre aussi une source d'erreur liée à l'imperfection du système oculomoteur. Cette erreur, dont les fondements sont basés sur les travaux de McConkie, Kerr, Reddix, Zola, et Jacobs (1989), diminue les saccades longues et allonge les saccades courtes. L'amplitude de la saccade est la somme de la distance prévue (c'est-à-dire la distance entre la fixation courante et le milieu du mot visé) et de deux erreurs, l'erreur aléatoire et l'erreur systématique. L'erreur systématique se base sur le fait qu'en lecture (en anglais du moins) les saccades de 7 caractères sont privilégiées. Cette erreur est modulée par la durée de fixation sur le lieu de départ : plus la fixation est longue plus la précision de la saccade augmente. L'erreur aléatoire augmente avec l'amplitude de la saccade. Ce premier modèle reste toujours très utilisé même si d'autres alternatives ont vu le jour.

SWIFT (GAG)

La principale différence entre *EZ-Reader* et SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) est que dans ce dernier plusieurs mots peuvent être vus et traités

en parallèle. C'est d'ailleurs en tant qu'alternative au principe de déplacement séquentiel de l'attention que SWIFT a vu le jour.

Les principes de ce modèle sont les suivants :

- Distribution spatiale du traitement dans un champ d'activation : la sélection du mot cible est due à une compétition entre les mots selon leur activation. Cette activation peut être vue comme une carte de saillance des mots, elle est représentée en gris sur la figure 1.6 page suivante.
- Séparation des voies spatiale et temporelle : le problème d'identification du moment où la programmation de la saccade commence et le problème de la localisation de la cible de la saccade sont découplés.
- Génération de la saccade autonome et inhibition fovéale : la durée de fixation dépend du temps de programmation de la saccade et est modulée par un processus d'inhibition fovéale qui augmente le temps d'inspection pour les mots compliqués.
- Programmation de la saccade en deux temps : durant la phase labile, la saccade est programmée, mais une nouvelle programmation peut être mise en place, qui annule la première. Durant la phase non labile, la cible est sélectionnée et la saccade ne peut plus être annulée.
- Erreur systématique et aléatoire dans l'amplitude de la saccade : en conséquence des saccades mal guidées peuvent entraîner des fixations mal placées.
- Corrections des fixations mal placées : lorsque les saccades « atterrissent » sur des mots non voulus, une nouvelle programmation démarre immédiatement.
- Modulation de la latence saccadique : la latence de la saccade, c'est-à-dire le temps mis pour initier la saccade, dépend de l'amplitude de celle-ci.

À partir d'une phrase ce modèle simule donc sa lecture en répondant lui aussi aux questions *où* et *quand*. Sur l'exemple figure 1.6 page suivante nous voyons que ce trajet comporte des mouvements typiques de la lecture : tous les mots ne sont pas fixés, certains mots reçoivent plus d'une fixation et des saccades reviennent dans le texte (saccades régressives). Les fixations sont en noir et l'aire d'activation en gris.

EMMA (SAS)

EMMA (Eye Movements and Movement of Attention) est un modèle développé par Salvucci (2001) qui simule les mouvements des yeux et l'encodage visuel. Il est associé à l'architecture ACT-R qui gère ici les processus cognitifs de plus haut niveau. L'avantage d'ACT-R est d'avoir des paramètres calibrés par ailleurs depuis plusieurs années de développement sur les processus bas niveaux. EMMA est basé sur le modèle *EZ-Reader* mais est aussi appliqué à des tâches autres que la lecture comme la résolution d'équations et la recherche visuelle.

L'architecture ACT-R Acronyme de *Adaptive Control of Thought-Rational*, ACT-R est une architecture cognitive symbolique. Il en existe d'autres telle que SOAR, développée par Newell, ou encore 3CAPS, DUAL, EPIC, Copycat ou Psi.

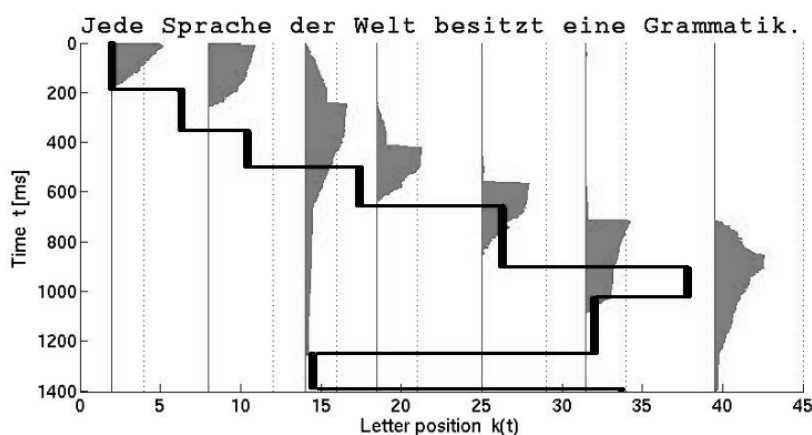


FIGURE 1.6 – Exemple de simulation du modèle SWIFT (Engbert et al., 2005). Le temps évolue selon l’axe vertical (de haut en bas). La position des fixations est indiquée par la ligne noire, sur l’axe horizontal, l’unité est la lettre. L’aire d’activation est indiquée par les zones grises.

Toutes ont un même but, mais des approches plus ou moins différentes. Le désir de créer ce type de modélisation est apparu avec Newell (1990) qui voulut avec une seule architecture (SOAR) rendre compte d’un comportement évolué générique, et expliquer comment tous les processus cognitifs, qu’ils soient ou non représentés sous forme de modules, fonctionnent ensemble pour produire une cognition cohérente, et les comportements relatifs. ACT-R fut développé par Anderson (1993); Anderson et al. (2004) et est fréquemment utilisé pour modéliser différents aspects du comportement humain, tel que la perception, l’attention, la mémoire, la résolution de problème, ou la prise de décision. ACT-R se divise en trois parties, la partie perceptuo-moteur, comportant quatre modules (visuel, langagier, auditif et moteur); la partie cognitive, avec deux mémoires, déclarative et procédurale; et la partie environnement, correspondant au monde extérieur.

Description du modèle Comme *EZ-Reader*, EMMA inclut une programmation du mouvement saccadique en deux temps (labile et non labile), et prend en compte la fréquence et l’excentricité pour implémenter le temps d’encodage. La grande différence entre les deux est que dans EMMA l’encodage et la programmation du mouvement se font simultanément. Cette différence permet au modèle de s’appliquer à d’autres tâches que la lecture. Le fait d’associer EMMA à une architecture (ici ACT-R) permet de le placer dans un cadre rigoureux fixant des paramètres et des contraintes comme le nombre d’items en mémoire, l’apprentissage des *chunks*¹, etc.

Quand un déplacement de l’attention est prévu vers un nouvel objet, EMMA commence à encoder l’objet pendant qu’un mouvement oculaire est préparé et (po-

1. Un chunk est un groupement d’éléments constituant une unité d’information. Par exemple un groupe de chiffre, comme une date, un numéro de téléphone; un groupe de lettres comme un mot, un acronyme; ou encore un groupe de mots, comme une phrase, une proposition (Miller, 1956)

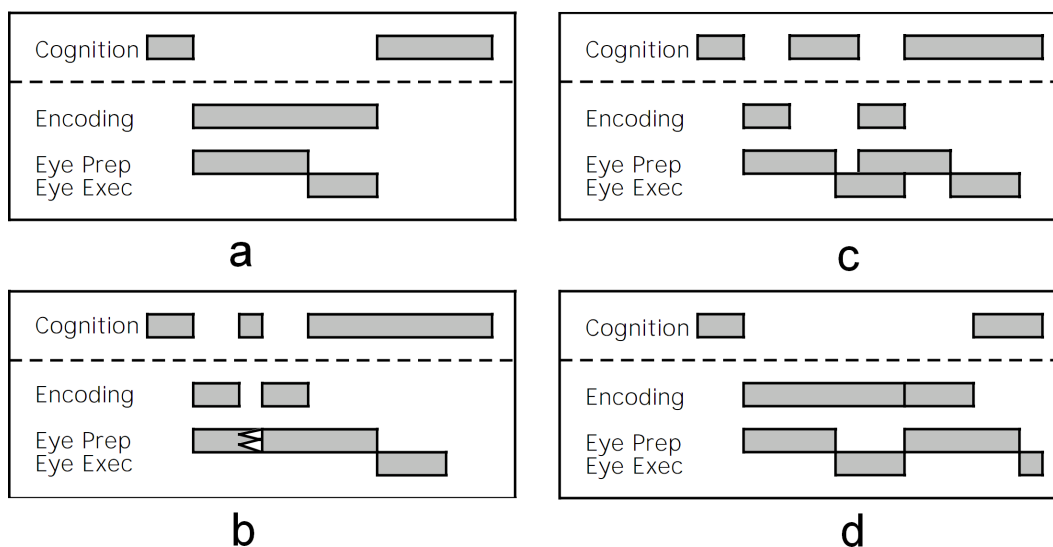


FIGURE 1.7 – Différents exemples de simulation du modèle EMMA (Salvucci, 2001), l'axe horizontal correspond au temps. (a) cas où l'encodage et le mouvement de l'œil sont simultanés, (b) un déplacement de l'attention annule l'étape labile entraînant l'omission de la fixation et la préparation d'un nouveau mouvement (c) un changement d'attention durant l'étape non labile ne perturbe pas le mouvement oculaire qui est exécuté mais un autre mouvement lui succède, (d) l'encodage continue après le premier mouvement oculaire induisant une refixation de l'objet. La ligne pointillée indique la séparation entre les processus cognitifs propres à l'architecture dans laquelle EMMA est déployé (ACT-R) et EMMA.

tentiellement) exécuté. Dans ce modèle le déplacement de l'attention et les mouvements oculaires sont deux processus distincts. Différentes simulations du modèle (figure 1.7) permettent de voir les deux temps de la programmation de la saccade (labile et non labile) et l'implication des processus cognitifs. Dans ce modèle les interactions se font entre les processus cognitifs (Cognition), l'encodage visuel (Vision), la préparation du mouvement oculaire (Eye-prep) et l'exécution du mouvement oculaire (Eye-Exec). Dans le cas le plus simple (a) l'encodage et le mouvement requièrent le même temps. Le module cognitif cède le contrôle à l'encodage visuel et le récupère quand l'encodage est terminé. Si l'encodage est terminé mais que le processus cognitif sollicite un nouveau déplacement de l'attention avant la fin du mouvement de l'œil, deux cas se distinguent. Si ce déplacement de l'attention se passe lors de la préparation du mouvement oculaire, ce mouvement est annulé et un nouveau mouvement se prépare (b), si ce déplacement de l'attention arrive lors de l'exécution du mouvement oculaire l'exécution continue jusqu'à être finalisée tandis qu'un nouveau mouvement oculaire commence (c). Enfin si le mouvement de l'œil est terminé avant la fin de l'encodage visuel, l'encodage continue et un nouveau mouvement se prépare (d), néanmoins comme le mouvement de l'œil a (certainement) mis l'objet dans la zone fovéale l'encodage est accéléré.

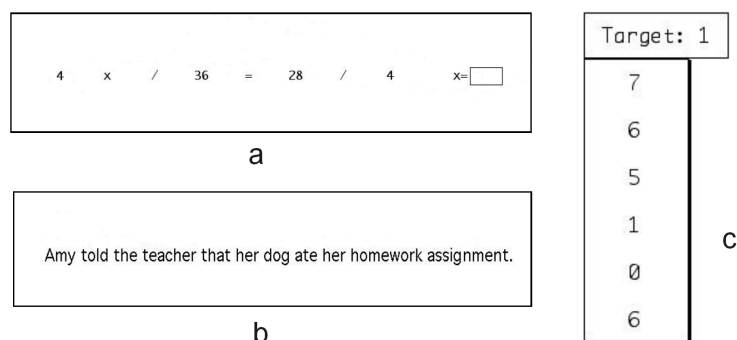


FIGURE 1.8 – Exemples d’applications du modèle EMMA (Salvucci, 2001). (a) résolution d’équation, (b) lecture linéaire, (c) recherche d’information visuelle dans une liste.

Validation du modèle EMMA a été testé sur trois tâches différentes : en lecture, en résolution d’équation et en recherche visuelle. Ces trois applications sont présentées figure 1.8. Les variables utilisées lors de la comparaison sont les temps de réalisation de la tâche ainsi que le nombre de fixations. Les comparaisons entre les performances du modèle et celles de participants montrent que ce modèle simule de manière cognitivement plausible les trajets oculaires des participants sur ces tâches (Salvucci, 2001).

Synthèse des modèles de lecture

Une comparaison des deux premiers modèles nous montre que plusieurs principes se retrouvent : la difficulté lexicale liée à la fréquence du mot et sa prédictibilité, le concept de traitement lexical en deux temps et la programmation de la saccade à deux niveaux, tout comme le concept d’erreur saccadique systématique et aléatoire. Ces modèles ne sont pas intégrés dans une architecture cognitive plus large. Le dernier modèle de lecture que nous avons vu, proche de *EZ-Reader* s’intègre lui à l’architecture ACT-R.

Ces modèles ont été comparés par Reichle, Rayner, et Pollatsek (2003), qui dressent un tableau synthétisant leurs aptitudes à reproduire les phénomènes observés en lecture (voir tableau 1.1 page suivante).

Dans ce tableau nous pouvons voir que *EZ-Reader* est celui qui est le plus proche des comportements observés. Ce tableau ne contient pas une liste exhaustive de tous les comportements rencontrés dans une tâche de lecture, mais il rassemble néanmoins des bases nécessaires pour valider la qualité d’un modèle de lecture.

Après avoir vu ces trois modèles de lecture, nous voyons bien que la modélisation de la lecture au niveau des fixations n’est pas triviale. Dans le cas de page Web, même si nous pouvons avoir recours à de la lecture linéaire, ce n’est pas le seul moyen de collecter des informations, celles-ci étant distribuées dans le document sous forme de paragraphes, de titres, de boutons, de listes, de menus.

TABLE 1.1 – Comparaison des modèles *EZ-Reader*, SWIFT et EMMA (Reichle et al., 2003).

	<i>EZ-Reader</i>	SWIFT	EMMA
Type de modèle	SAS	GAG	SAS
Distributions des positions des fixations	oui	non	non
Plage d'erreur systématique	oui	non	non
Mesure basée sur les mots	oui	limité	oui
Effets de fréquence	oui	oui	oui
Prévisualisation parafovéale	oui	limité	limité
Effets de débordement	oui	non	non
Coûts d'omission	oui	oui	non
Effets de prédictibilité	oui	oui	non

Un autre mécanisme de parcours entre les éléments est donc nécessaire afin d'effectuer une recherche dans un document structuré en blocs. Nous nous placerons donc dans le cadre de la recherche d'information sur le web afin d'évoquer deux des principaux modèles existants, se basant sur deux architectures cognitives différentes. Ces deux modèles simulent la navigation entre pages web et au sein d'une même page.

1.3.2 SNIF-ACT

SNIF-ACT (*Scent-base Navigation and Information Foraging in the ACT architecture*) est un modèle qui simule le comportement des utilisateurs lors d'une tâche de recherche d'information sur le Web (Pirolli & Fu, 2003; Fu & Pirolli, 2007). SNIF-ACT choisit les actions qu'il va effectuer sur la base d'une mesure d'information : l'*Information Scent*, calculée à partir des informations de la page et du but de l'utilisateur. Les prédictions du modèle ne sont pas des fixations, ni des trajets oculomoteurs, mais les actions réalisées, par exemple les clics sur les liens ou les changements de pages. SNIF-ACT est implémenté dans l'architecture ACT-R. Il se situe au niveau de la partie cognitive d'ACT-R, c'est-à-dire sa mémoire procédurale et sa mémoire déclarative.

Structure de SNIF-ACT

L'architecture de SNIF-ACT est représentée figure 1.9 page suivante. Dans la mémoire déclarative se trouvent les connaissances sous forme de *chunks*, tel que le contenu des liens hypertextes, ou les fonctionnalités des boutons du navigateur. Les *chunks* ont chacun une valeur d'activation qui reflète leur pertinence par rapport au focus attentionnel courant.

Dans la mémoire procédurale se trouvent des règles de procédure, sous forme de couples condition-action. Si les conditions sont réunies, c'est-à-dire les *chunks* activés

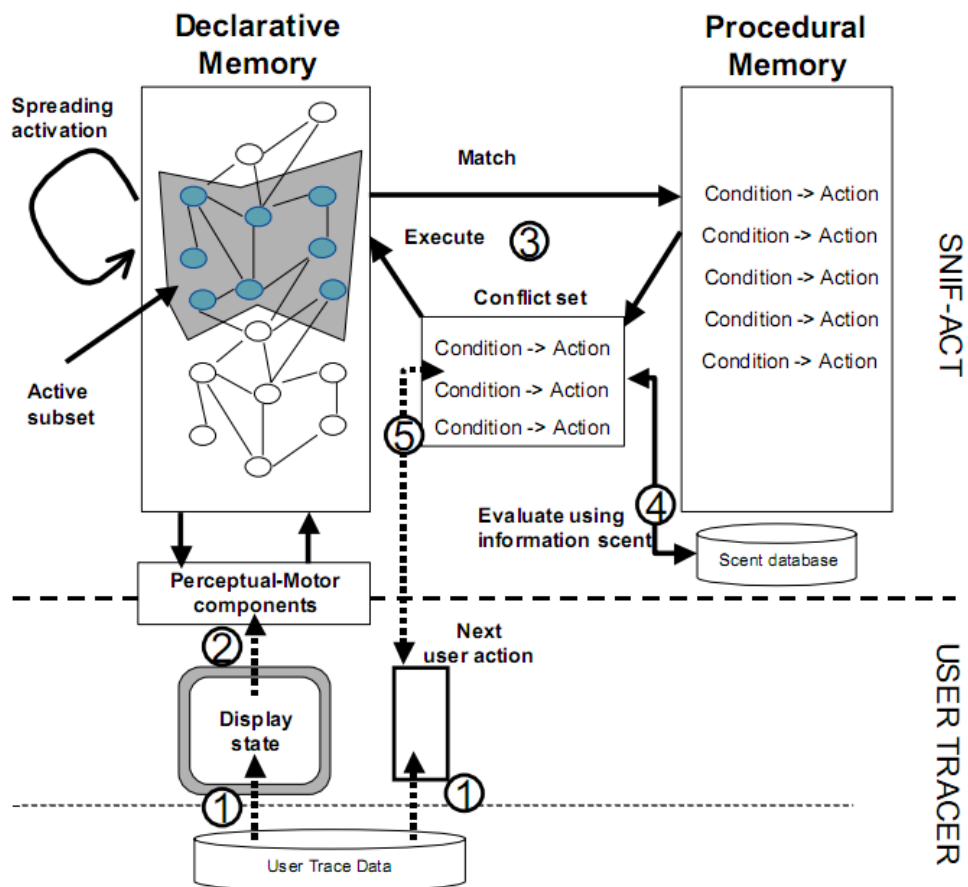


FIGURE 1.9 – Structure du modèle SNIF-ACT, les nombres indiquent l'ordre des étapes d'un cycle lors de la simulation (Pirolli & Fu, 2003).

TABLE 1.2 – Exemples de productions de SNIF-ACT

Use-Search-Engine : IF the goal is Goal*Start-Next-Patch & there is a task description & there is a browser & there is a search engine name in memory & the browser is not at the search engine THEN Set and push a subgoal Goal*Use-Search-Engine to the goal stack
Start-Process-Page : IF the goal is Goal*Start-Next-Patch & there is a task description & there is a browser & the browser is on an unprocessed page THEN Set and push a subgoal Goal*Process-Page to the goal stack

en mémoire déclarative alors l'action est effectuée (voir les exemples tableau 1.2). À chaque étape une production est sélectionnée, et s'il y a plus d'une production correspondante, un mécanisme de résolution des conflits est appelé pour choisir celle qui sera exécutée. Ce mécanisme de résolution des conflits est basé sur sa fonction d'utilité (*l'Information Scent*).

Afin de faire des choix pertinents lors d'une recherche d'information sur le Web, plusieurs éléments entrent en compte, les liens, les images, le texte. Ces informations sont prises en compte localement pour calculer cette valeur d'utilité. Deux types de *chunks* vont ainsi être activés, ceux liés au but de l'utilisateur et ceux liés aux éléments de la page. Ces *chunks* sont activés par un mécanisme de diffusion de l'activité (*spreading activation*) par leurs liens d'associations. Ces associations entre mots dans la mémoire sont reliées aux probabilités d'occurrence et co-occurrence des mots. Ces probabilités sont calculées à partir du corpus *Tipster*² et du Web.

Ce modèle a été comparé à des données empiriques sur deux tâches de recherche d'information.

Évaluation de SNIF-ACT

La première version de SNIF-ACT (Pirolli & Fu, 2003) a été comparée à des données expérimentales sur deux tâches : trouver un site où acheter des posters du film *Fourmiz* (*Antz Task*) et trouver un site avec la date d'une pièce de théâtre, ainsi qu'une photo de la troupe (*City Task*). Les comparaisons ont porté sur les variables (actions) suivantes : quels sont les liens cliqués sur une page et à quel

2. *Tipster* est un corpus de textes d'environ 450 millions de mots principalement issus de la presse écrite américaine (*Associated Press*, *WallStreet journal*, *Ziff-Davis Publishing*, etc) sur une période de 1986 à 1992.

moment les pages sont abandonnées ? Les résultats ont montré que les liens choisis par les participants avaient une valeur informative (*Information Scent*) calculée par SNIF-ACT élevée. Cette mesure est aussi prédictive du moment où les utilisateurs quittent une page. Chaque page est représentée par une valeur informative moyenne. Lorsque les utilisateurs quittent la page, la valeur calculée par le modèle est en effet plus faible que la valeur moyenne sur toutes les pages visitées par l'utilisateur.

Ce modèle a ensuite été comparé à un modèle de position, qui décide de sélectionner un lien en se reposant uniquement sur la position des liens sur une page Web. Les données simulées par SNIF-ACT 1.0 sont plus proches des données expérimentales que celles issues de ce modèle de position, mais les auteurs ont considéré que la sélection des liens dépendait sûrement de l'interaction dynamique entre l'*Information Scent* et la position des liens sur une page Web. Ils ont donc amélioré leur modèle en incluant un mécanisme bayésien qui décide dynamiquement quel lien suivre et quand abandonner une page Web en prenant en compte l'ordre des liens. SNIF-ACT 2.0 (Fu & Pirolli, 2007) est donc plus souple et mieux adapté à la dynamique des interactions entre l'utilisateur et les différents sites Web. Les simulations de ce nouveau modèle ont été comparées à un ensemble de données recueillies dans une étude impliquant 74 utilisateurs cherchant des informations sur deux sites Web : help.yahoo.com (la section d'aide de Yahoo!) et parcweb.parc.com (l'intranet d'une entreprise). Les résultats ont été meilleurs pour ce modèle que pour SNIF-ACT 1.0 et que pour le modèle de position.

Ce modèle est donc intéressant du point de vue de la navigation entre les pages web. Avoir un outil de calcul des similarités entre les informations perçues, les buts et objectifs de l'utilisateur et les connaissances a priori permet de simuler les comportements des utilisateurs de pages Web en recherche d'information.

Dans le modèle que nous allons décrire maintenant cette prise en compte de l'information sémantique est essentielle. Ce modèle se base sur l'architecture cognitive développée par Kintsch (1998) de Construction-Intégration (C-I) pour la compréhension de textes, contrairement à SNIF-ACT basé, lui, sur la théorie d'*Information Foraging* et ACT-R.

1.3.3 CoLiDeS

CoLiDes, acronyme de *Comprehension-based Linked model of Deliberate Search* a été développé par Kitajima et al. (2000) ; Kitajima, Blackmon, et Polson (2005). Il s'agit d'un modèle centré sur quatre processus cognitifs : analyse de la page (*parsing*), focalisation (*focusing on*), compréhension (*comprehension*) et sélection (*selection*). Ces processus sont répartis en deux phases, une phase d'attention, le premier cycle de Construction-Intégration et une phase d'action, le second cycle de Construction-Intégration. L'architecture de Construction-Intégration décrite par Kintsch (1998) fait une séparation sensiblement différente des phases, à savoir une première phase de récupération des informations, et une seconde phase d'intégration de l'item courant avec le reste.

Principe de CoLiDeS

Lors d'une recherche d'information CoLiDeS commence par prendre en compte le but de la recherche, et pour cela il se crée des sous-but. Il existe deux types de sous-but, ceux de contenu (quelles informations chercher?) et ceux de navigation (comment trouver une page Web?). Nous retrouvons cette première étape sur la figure 1.10 page 29. Sur cette figure le sous-but de contenu est de trouver des informations sur le diabète de type 2, et le sous-but de navigation est d'utiliser la barre de navigation du site (à gauche de la page Web). Une fois ses buts définis, le modèle entre dans la première phase, la phase d'attention.

Dans cette première phase, la page est tout d'abord séparée en sous-régions à partir des connaissances en mémoire à long terme et des informations visuelles disponibles. Ce modèle se base sur le fait que les utilisateurs prennent en compte les textes qui se trouvent sur la page, mais aussi les logos, les menus, les liens, ou tout autre composant de l'interface pour créer une division de la page en sous-régions. Sur la figure 1.10 nous voyons que la page est décomposée en logo, barre de navigation, fenêtre de recherche, derniers sujets, et publicité. Une sous-région est ensuite sélectionnée selon sa proximité avec les sous-but de contenu et de navigation. Nous passons alors dans la phase d'action.

Lors de cette seconde phase, les objets de la région sélectionnée sont précisés, le modèle regarde plus en détails les informations sur chaque élément, du point de vue sémantique et fonctionnel. Quelques objets, les *meilleurs*, sont sélectionnés sur différents critères. La notion fondamentale de ce modèle est cette sélection des *meilleurs* objets, l'équivalent de l'*Information Scent* dans SNIF-ACT. Dans CoLiDeS il y a cinq facteurs qui entrent en jeu pour définir le lien entre un objet et le but de l'utilisateur. Le modèle utilise l'analyse de la sémantique latente (LSA, *Latent Semantic Analysis*) pour trois de ces facteurs comme mesure de la similarité sémantique entre les éléments³.

Les cinq facteurs sont les suivants :

- le degré de similarité sémantique entre le but de l'utilisateur et le titre/liens (LSA) ;
- s'il y a un niveau adéquat d'information a priori pertinente pour pouvoir préciser l'élément.
- si un des mots de l'élément a une très basse fréquence d'utilisation (LSA) ;
- la fréquence avec laquelle l'utilisateur a rencontré les élément auparavant, quelqu'un qui a déjà utilisé le moteur de recherche d'un site va avoir une probabilité plus élevée d'y retourner, plutôt que de passer par une navigation par menus ;
- s'il y a une correspondance partielle ou totale entre le but de l'utilisateur et un élément de l'écran.

La combinaison de ces cinq facteurs donne à chaque élément une valeur.

3. LSA est un modèle cognitif de représentation du sens des mots. Cette méthode estime les relations sémantiques entre mot ou groupe de mots, elle est basée sur une analyse statistique de large corpus (Landauer & Dumais, 1997). LSA est décrit dans la section 2.4.2 page 40.

Une fois ces éléments traités, les actions qui seront effectuées sont sélectionnées à leur tour, puis réalisées. Par exemple dans le cas d'une recherche sur internet, les actions associées avec les liens sont pointer (*Point*) et cliquer (*Click*) (voir sur l'exemple de réalisation figure 1.10 page ci-contre).

L'intérêt de ce modèle est sa capacité à prendre en compte les connaissances a priori de l'utilisateur avec l'utilisation de LSA. De plus, CoLiDeS a la capacité d'apprendre les séquences d'actions précédentes et leurs conséquences pour les réutiliser dans des recherches futures.

Ce modèle est néanmoins limité car il n'est pas entièrement automatique, contrairement à SNIF-ACT. En effet, la page est découpée en régions au préalable. De plus la sélection du lien n'entraîne pas de navigation car le modèle ne fonctionne que sur une page unique.

1.3.4 Complémentarité de CoLiDeS et SNIF-ACT

Ces deux modèles sont complémentaires et nous donnent plusieurs éléments clés sur la navigation sur le Web et les différents éléments à prendre en compte pour simuler une recherche d'information. Kitajima, Polson, et Blackmon (2007) en font une comparaison intéressante.

Les auteurs mettent en avant les points forts des deux approches, notamment le fait de confronter deux théories, l'*Information Foraging* d'un côté et la *Construction-Integration* de l'autre. Ces deux modèles simulent le mécanisme qui permet l'évaluation des liens entre les pages en utilisant une mesure de distance entre le lien et le but de la recherche dans un espace sémantique. Dans le cas de SNIF-ACT les pages sont vues comme des entités informatives alors que pour CoLiDeS une même page se divise en plusieurs patches (sous-régions), ces modèles s'intéressent donc à des niveaux différents de traitements de l'information. Ils nous montrent bien l'importance de la prise en compte des informations sémantiques dans la navigation ainsi que la complexité de la recherche d'information dans des documents type page Web. En revanche ils ne prennent pas en compte les mécanismes de lecture : la simulation ne descend pas au niveau plus fin des saccades et des fixations.

1.4 Problématique

Notre objectif est de comprendre et de modéliser quels sont les processus cognitifs nécessaires à la réalisation d'une tâche de recherche d'information ainsi que leurs interactions. Ce modèle s'intéresse au trajet oculomoteur d'un utilisateur en quête d'information sur un sujet précis. Les fixations et saccades simulées par le modèle se situent à plusieurs niveaux. Au niveau des transitions entre éléments, comme dans CoLiDeS et à un niveau plus précis dans un paragraphe avec des mécanismes de lecture qui vont entrer en jeu. Les différents modèles que nous avons vus nous permettent de nous positionner par rapport à cet état de l'art.

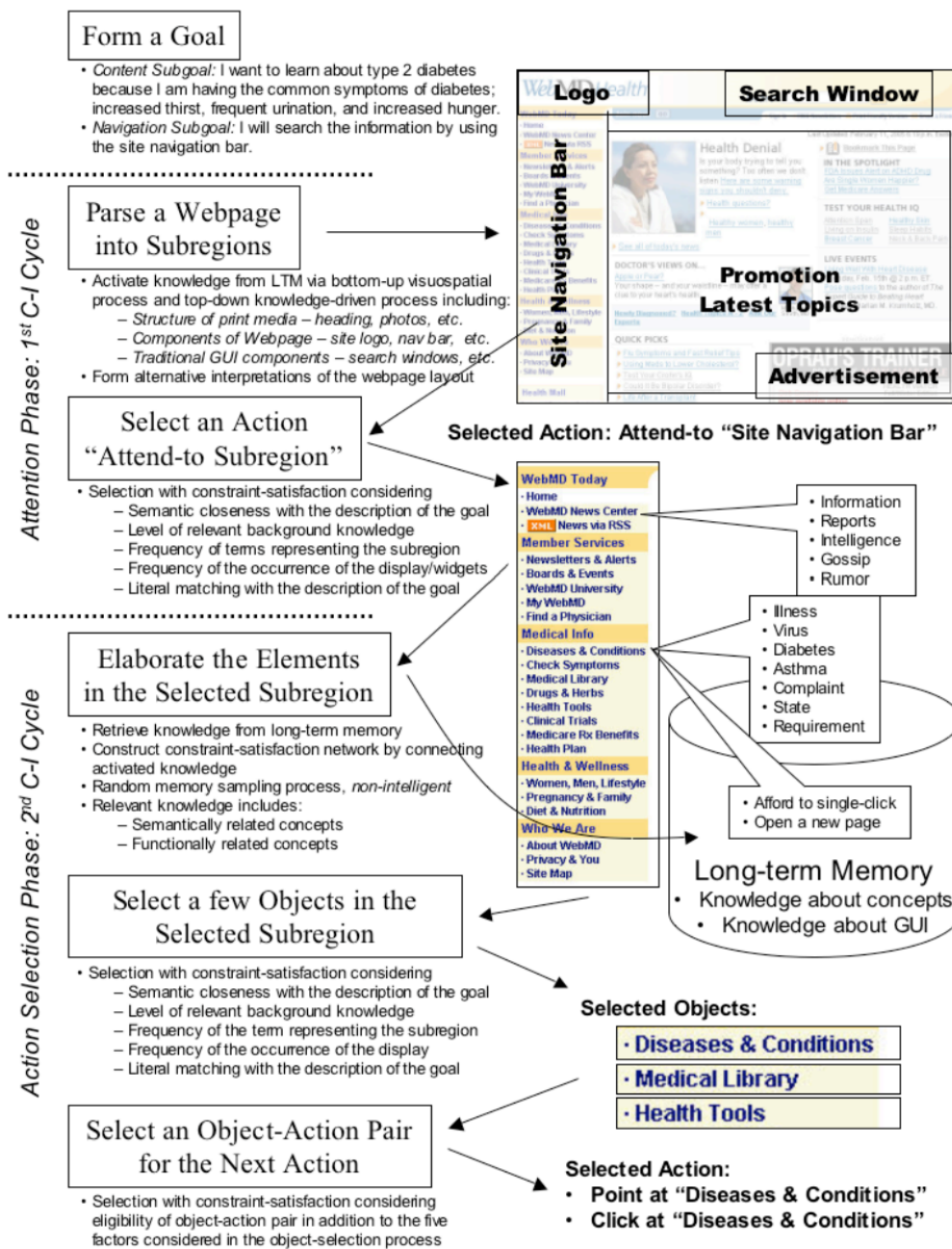


FIGURE 1.10 – Simulation de CoLiDeS pour une recherche sur le diabète de type 2. L’action sélectionnée est de cliquer sur le lien *Diseases & Conditions*.

Nous nous plaçons dans le cadre de la simulation cognitive de trajet oculomoteur lors d'une recherche d'information dans des documents comportant des informations visuelles et sémantiques. Différents types d'information doivent être pris en considération, le texte, et sa valeur sémantique, les images du document, les mises en forme, les mots mis en valeur, les titres, etc. Aucun des modèles précédents ne s'attaque à un problème similaire même si tous partagent une partie de notre objectif. Nous allons donc prendre en compte ces diverses contributions pour élaborer notre modèle.

2

Présentation générale du modèle

Notre modèle se base sur une intégration des processus qui entrent en jeu lors d'une tâche de recherche d'information et ce, dans le but de prédire une séquence de fixations proche de celles des humains.

Par exemple lors d'une recherche sur une page Web, un utilisateur peut faire le trajet figure 2.1 page suivante (enregistré par T. Baccino). Ce trajet a été généré alors que l'utilisateur était à la recherche d'une information sur la *jurisprudence*. Les éléments saillants visuellement l'ont attirés, par exemple les titres qui sont plus gros, et en couleur, mais les titres sont aussi porteurs d'informations sémantiques, ils sont en général le reflet du paragraphe auquel ils sont liés. D'autres éléments saillants ont en revanche été négligés, les images, le titre de la page *Bienvenue sur le site de l'ANIL*. Nous pouvons aussi émettre l'hypothèse qu'un mécanisme mnésique a permis à l'utilisateur de ne pas revenir sur les parties déjà traitées. Enfin une stratégie globale semble avoir été mise en place, afin de faire un parcours de la page systématique (de haut en bas et de gauche à droite) et peu coûteux jusqu'au but.

Nous avons cherché à préciser et modéliser ces intuitions en nous basant sur différentes études en vision, mémoire, sémantique et recherche d'information. Notre modèle est donc une combinaison de ces mécanismes, qui vise à simuler les trajets oculomoteurs humains. Le modèle ne s'applique pas sur de réelles pages web pour des raisons de contrôle expérimental, mais dans l'avenir ce modèle pourra très certainement être appliqué à des documents réels.

Nous avons choisi différentes applications grâce auxquelles nous allons tester et améliorer notre modèle. Avant d'entrer en détail dans sa description, nous allons



FIGURE 2.1 – Exemple de trajet oculomoteur d'un utilisateur à la recherche d'une information sur une page Web.

décrire ces applications. Nous verrons ensuite l'architecture générale du modèle, puis pour chacun des processus la façon dont ceux-ci peuvent être implémentés en se basant sur les modèles théoriques issus de recherches en sciences cognitives.

2.1 Application du modèle

Le modèle que nous décrivons dans la section suivante s'applique sur plusieurs tâches de recherche d'information. Ces tâches sont de plus en plus complexes afin de tester et valider le modèle pas à pas. Nous allons voir ici les différents types de matériel auquel le modèle peut s'appliquer.

La première tâche est la recherche d'un mot parmi plusieurs sur un écran. L'objectif est de répondre à une question le plus efficacement possible (avec le meilleur mot) la cible n'est donc pas connue. Les mots ont une valeur sémantique et visuelle. Cette expérience a été reproduite avec différentes caractéristiques visuelles. Un exemple se trouve figure 2.2 page ci-contre, la question est : *Quel est l'instrument de musique le plus petit ?*

La seconde tâche du modèle est une recherche d'information dans un document plus complexe, avec des éléments plus variés, textes et photographies. Dans ce cas la consigne est de rechercher le paragraphe qui est le plus proche d'un thème donné, par exemple : *Quel est le bloc le plus associé au thème Chef de la Russie ?* (voir figure 2.3 page suivante).

Enfin la dernière tâche du modèle est une recherche d'information proche de la seconde, les thèmes et les textes sont les mêmes mais la disposition dans la page

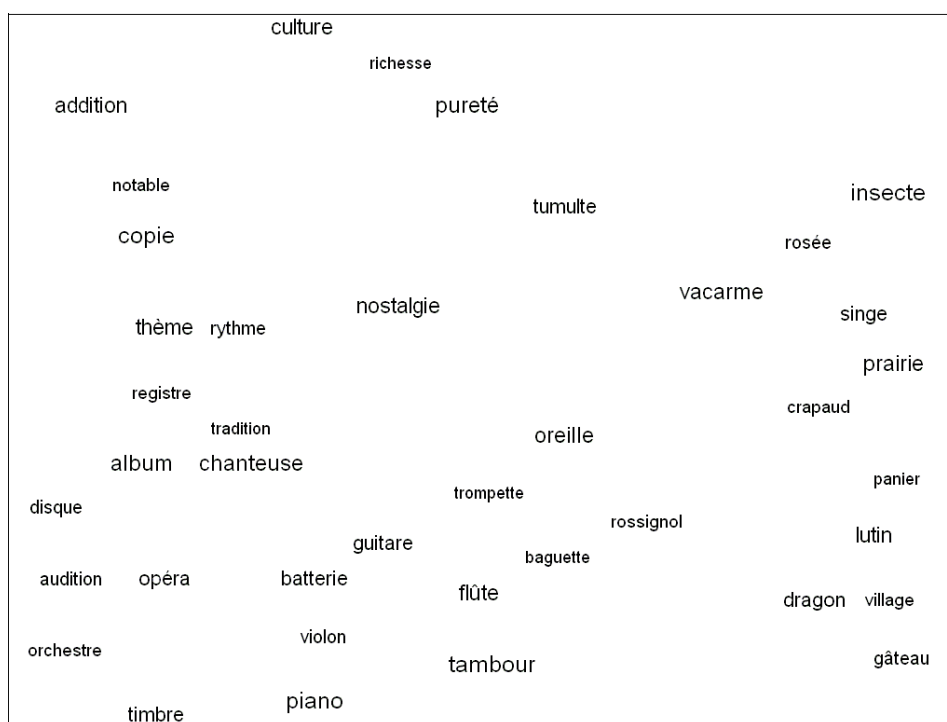


FIGURE 2.2 – Exemple d’application. Objectif de la recherche, trouver l’instrument de musique le plus petit.

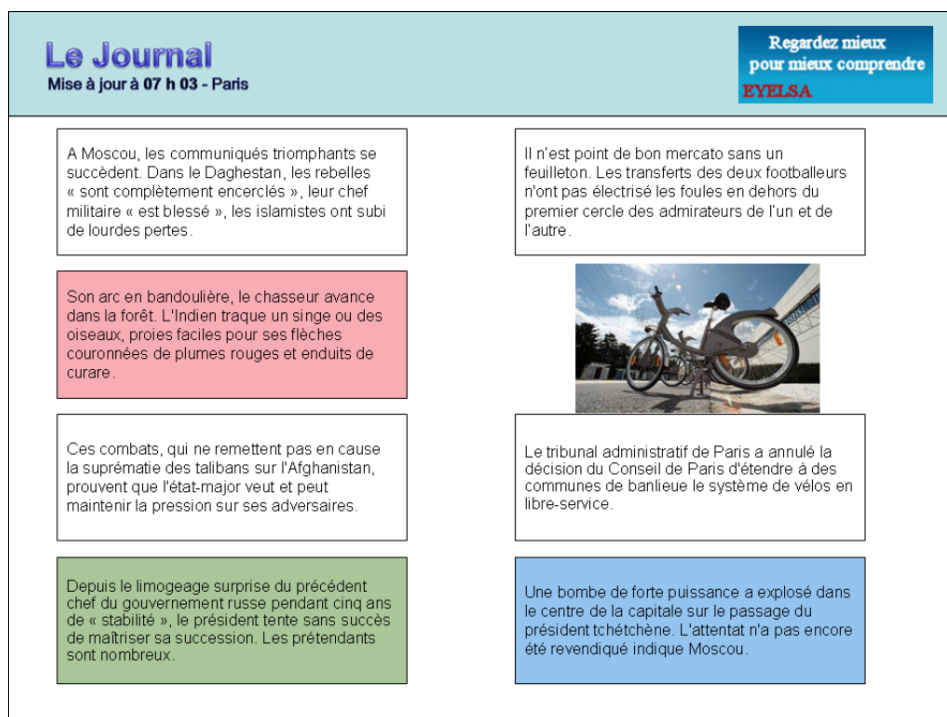


FIGURE 2.3 – Exemple d’application. Objectif de la recherche, trouver le paragraphe le plus proche du thème *Chef de la Russie*.

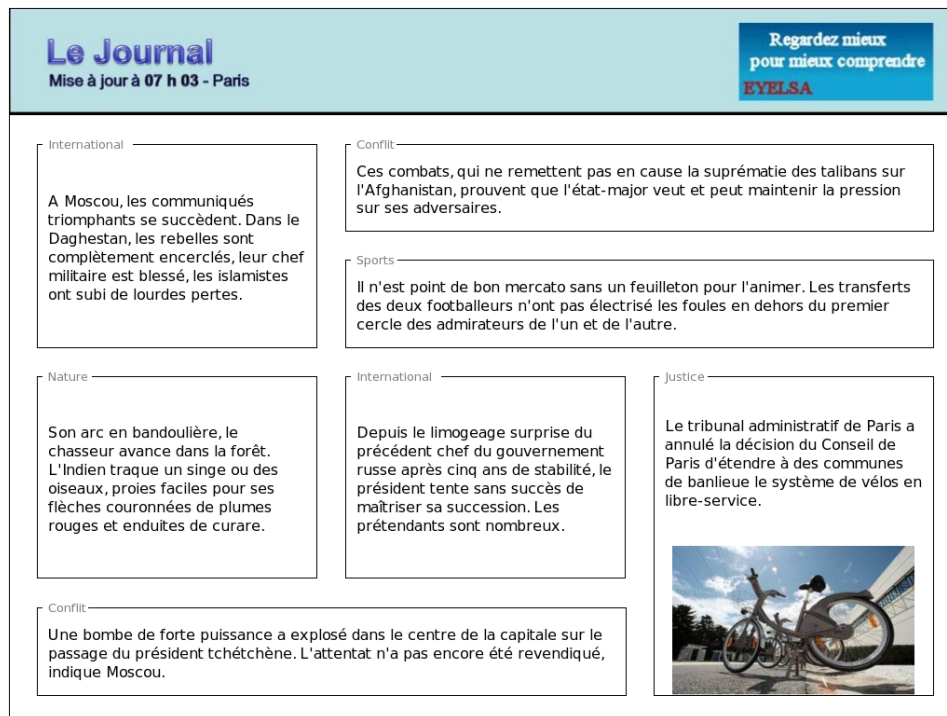


FIGURE 2.4 – Exemple d'application. Objectif de la recherche, trouver le paragraphe le plus proche du thème *Chef de la Russie*.

change, et des titres sont ajoutés en relation avec les paragraphes dans une modalité (voir figure 2.4).

Ces choix d'applications, comprenant à chaque fois des nouvelles informations à traiter vont nous permettre d'améliorer et de tester l'efficacité et la robustesse du modèle.

2.2 Architecture

L'idée principale est la prise en compte de plusieurs sources d'informations et leur intégration pour générer un trajet simulant le déplacement de l'attention lors d'une tâche de recherche d'information. Ces informations sont de deux types : des informations visuelles, et des informations sémantiques. Nous avons cherché quels étaient les processus mis en jeu dans une tâche de recherche d'information et nous en avons identifié trois. Le premier est un processus visuel, le second est un processus sémantique et le dernier est un processus mnésique. Cette séparation en trois processus, interagissant entre eux, est une simplification qui nous a permis d'implémenter un modèle et de tester nos hypothèses. La partie visuelle nous a paru nécessaire au traitement des éléments visuels de la page, ainsi qu'à la prise en compte de caractéristiques physiques de l'œil. La partie sémantique, elle, permet à l'utilisateur, et au modèle, d'avoir accès à des connaissances sémantiques afin de traiter, lire et comprendre les informations textuelles présentes dans la page. Cette distinction n'est

pas toujours aussi franche, ainsi un élément en rouge est certes saillant visuellement, mais la couleur rouge a aussi un sens car elle est souvent utilisée pour signifier une importance, elle a donc une valeur sémantique, mais dans une volonté de simplification nous avons séparés les informations sémantiques et visuelles. La partie mnésique permet elle d'avoir un suivi entre les actions en ayant une trace en mémoire des éléments déjà visités. Là encore comme nous le verrons par la suite, il s'agit d'une simplification car notre modélisation de la mémoire se limite à une mémorisation des éléments dans leur globalité, sans faire de distinction entre les différentes formes de mémoire décrites dans la littérature (Baddeley & Hitch, 1974).

De manière générale, l'architecture du modèle est représentée figure 2.5 page suivante. Sur cette figure nous voyons les trois processus représentés par trois cartes. Ces trois cartes (visuelle, sémantique et mnésique) contiennent le poids de chaque élément pour le processus qu'elles représentent à chaque instant. Pour l'intégration des trois processus, nous avons choisi de faire une somme pondérée. Nous n'avons pas d'a priori pour tel ou tel type d'intégration, une somme nous a donc parue la solution la plus simple. Les entrées du modèle (en haut sur ce schéma) sont le but de l'utilisateur (ce qu'il cherche) et la page sur laquelle il va effectuer cette tâche de recherche d'information. La carte sémantique est construite à partir du but de l'utilisateur, des informations sémantiques de la page et de connaissances sémantiques. La carte visuelle, construite à partir des informations visuelles de la page, est construite en deux étapes, tout d'abord une carte de saillance qui est modulée par un filtre rétinien, que nous décrirons ultérieurement. Enfin la carte mnésique est elle construite à partir des éléments de la page et à partir des actions précédentes mémorisées en mémoire de travail. À partir de la carte finale, c'est-à-dire de la somme des trois cartes à un instant t , la localisation de la prochaine fixation est déterminée, une saccade est générée, suivie d'une fixation à l'endroit choisi. Les cartes sont ensuite régénérées conditionnellement à la localisation de la nouvelle fixation et un nouvel emplacement est choisi, etc. Ce modèle est donc dynamique dans le sens où chaque carte est régénérée à chaque étape. La sortie du modèle est un trajet oculomoteur sur la page. Ces processus peuvent agir à plusieurs niveaux. Nous distinguerons ici les niveaux global et local.

Au niveau global, le processus visuel va extraire, si possible, l'aspect général de la page, ce qui lui permettra de se faire une représentation globale de la page pour, par la suite, planifier son parcours. Le processus sémantique prend en compte le fait qu'en globalité les éléments sont arrangés sémantiquement dans l'espace, c'est-à-dire qu'une proximité spatiale correspond à une proximité sémantique. Enfin le processus mnésique prend en compte la globalité du parcours déjà effectué.

Au niveau local, le processus visuel va prendre en compte localement les caractéristiques visuelles des éléments autour de la fixation considérée. Le processus sémantique, de la même manière, va considérer la similarité sémantique de l'élément fixé pour décider de rester ou non dans la zone fixée. De plus lors de la lecture de paragraphe ce processus va traiter localement les mots un à un pour prendre

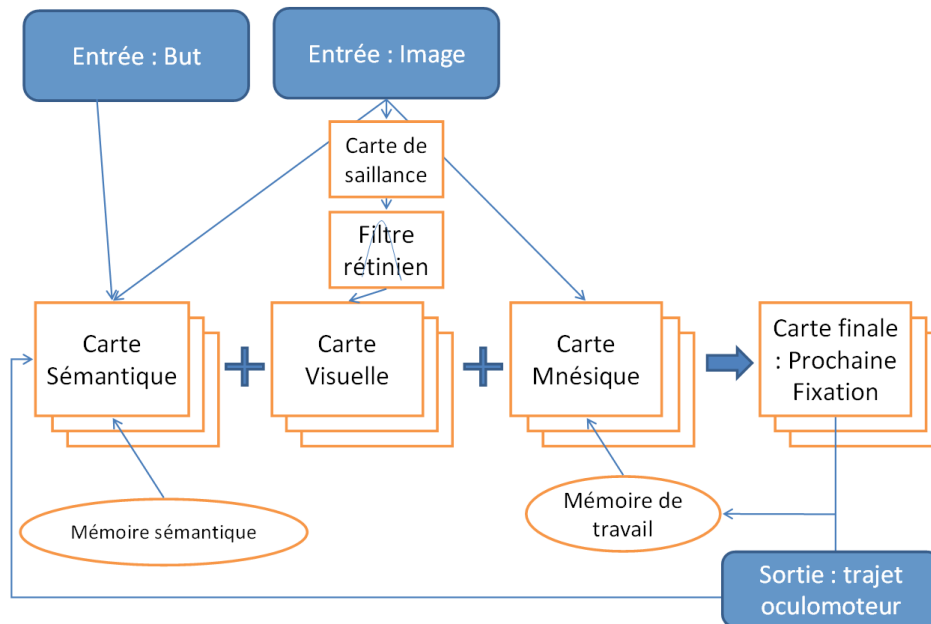


FIGURE 2.5 – Architecture du modèle.

une décision. Enfin le processus de mémoire, avec un mécanisme d'oubli donne plus d'importance à l'élément actuellement fixé.

Ces deux niveaux vont donc se retrouver tout au long de ce chapitre sur les différents processus et leurs modélisations. Nous allons maintenant voir plus en détail la génération de ces cartes en séparant les processus même s'ils restent fortement liés, notamment la sémantique et la mémoire.

2.3 Processus visuel

2.3.1 Éléments théoriques

Dans l'optique d'une modélisation de mouvements oculaires nous verrons tout d'abord ce que sont les mouvements oculaires et comment les utiliser comme observables du comportement humain. Nous verrons ensuite la notion d'acuité visuelle.

Mouvements oculaires

Lors d'une recherche d'information visuelle, ou lorsque tout simplement nous posons notre regard sur le monde qui nous entoure, nous effectuons des mouvements oculaires, reflets de notre activité cognitive. Ces mouvements oculaires sont de plus en plus utilisés dans la recherche comme données objectives pour essayer de mieux comprendre le mécanisme de déplacement de l'attention dans l'environnement. Des expérimentations avec enregistrement des mouvements oculaires ont permis des avancées dans la compréhension des processus cognitifs dans différents

domaines. Au niveau de la lecture, plusieurs études ont ainsi montré quels mots ou quelles parties des mots sont fixés, quels sont les taux de saccades régressives (qui reviennent sur un texte déjà vu), etc. Au niveau de la vision, citons entre autres les travaux de [Henderson et al. \(2007\)](#), qui montrent par une étude oculométrique que les modèles de cartes de saillance ne correspondent pas aux données expérimentales. En IHM (Interaction Homme Machine), les apports de l'utilisation de l'oculomètre pour recueillir des informations sur le comportement des utilisateurs sont là aussi conséquents ([Chalon et al., 2001](#)).

Ces enregistrements oculaires doivent être analysés pour nous apporter des informations pertinentes, et contrairement à ce que l'introspection pourrait nous laisser penser notre regard ne se déplace pas de manière continue mais est une succession de fixations, de saccades, et de micro-mouvements de moindres amplitudes.

Fixations, Saccades Les mouvements oculaires se divisent classiquement en des périodes de relative stabilité appelées fixations qui alternent avec des périodes de mouvements rapides appelés saccades. Notre regard se déplace constamment et nous faisons en moyenne 2 à 3 fixations par seconde. La durée des fixations est très variable. En lecture, les fixations durent en moyenne environ 250 ms (entre 50 et 500 ms typiquement). Elles sont en général un peu plus longues en recherche visuelle (275 ms) et en perception de scènes visuelles (330 ms). La durée des saccades est beaucoup plus brève. Une saccade de deux degrés¹ (huit caractères), typique en lecture, dure environ 30 ms. Une saccade plus longue de 5 degrés, typique en perception de scène, dure environ 40-50 ms.

D'autres mouvements existent, telles les micro-saccades. Leur étude est plus récente car les capturer avec un oculomètre demande une très grande finesse, qu'il était techniquement impossible d'atteindre il y a plusieurs années. Leur lien avec la charge cognitive est en débat, mais est évoqué dans certaines études ([Steinhauer, 2002](#)).

Acuité visuelle

Nous n'entrerons pas ici dans le détail de la physiologie de l'œil, mais un bref aperçu nous permettra de comprendre la notion d'acuité visuelle qui sera intégrée dans notre modèle. La partie de l'œil humain qui nous intéresse est la rétine, qui tapisse le fond de l'œil. La lumière traverse les différentes couches rétinienne pour arriver sur les photorécepteurs, qui sont de deux types : cônes et bâtonnets. Les cônes sont dédiés à la vision diurne et aux couleurs tandis que les bâtonnets sont eux dédiés à la vision nocturne et en niveaux de gris. La répartition des photorécepteurs n'est pas uniforme sur la rétine. Les bâtonnets sont beaucoup plus nombreux dans la rétine périphérique. Les cônes sont très majoritairement dans la fovéa, et leur densité décroît avec l'excentricité (voir figure 2.6 page suivante).

1. Un degré dans le champ visuel équivaut à environ 1cm sur un écran situé à 57 cm, ou l'ongle du pouce à bout de bras.

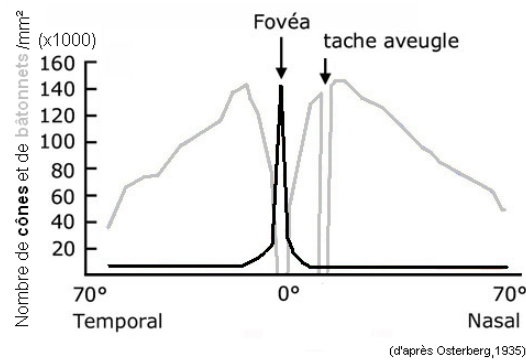


FIGURE 2.6 – Répartition des cônes (en noir) et des bâtonnets (en gris) sur la rétine en fonction de l'angle visuel. La fovéa est la région où la vision est la plus précise, grâce à sa forte densité de cônes.

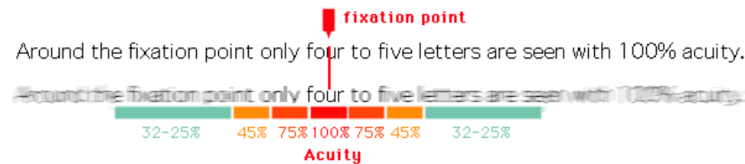


FIGURE 2.7 – Simulation de l'acuité visuelle en lecture (Hunziker, 2006). La fixation est sur le mot *four* et nous voyons bien la dégradation de la vision autour de ce point sur la seconde ligne.

Cette répartition produit une variation dans l'acuité visuelle qui est maximale au centre de la rétine (fovéa) et chute considérablement en s'éloignant du centre. Les fixations permettent de se placer sur les éléments dont nous voulons avoir une vision précise (exemple en lecture figure 2.7). La forme générale de la distribution de l'amplitude des saccades lors d'une tâche de recherche visuelle ou d'exploration de scène nous montre que les trajets oculomoteurs comportent beaucoup de saccades de faible amplitude et peu de longues saccades. Cette observation est en accord avec l'acuité visuelle car les saccades sont faites vers des éléments proches de la fovéa. Un bon modèle de simulation de trajets oculomoteurs doit donc prendre en compte l'acuité visuelle afin d'obtenir des distributions d'amplitude des saccades proches de celles observées empiriquement (Ho Phuoc, Guérin-Dugué, & Guyader, 2009). Au contraire, les modèles n'intégrant pas ce mécanisme de filtrage rétinien considèrent que toute l'image a une acuité maximale et la distribution de l'amplitude des saccades est alors proche de l'uniforme.

Pour une bonne simulation des parcours oculomoteurs, il nous est donc apparu nécessaire de prendre en compte cette notion d'acuité.

2.3.2 Pistes de modélisation

Grâce à l'observation des trajets produits par les participants dans les différentes applications, nous avons mis en évidence deux types de comportement au niveau visuel. Le premier, que nous appellerons réactif ou local permet de rendre compte des contraintes locales lors de la pondération des éléments dans la carte visuelle. Ces contraintes prennent en compte plusieurs propriétés du système visuel : l'acuité visuelle, correspondant au fait que la qualité de la vision se dégrade rapidement en s'éloignant du point de fixation, et la saillance visuelle indiquant que certains éléments ont des traits saillants attirant plus fortement l'attention que d'autres. Lorsqu'une fixation est sur un élément, le choix de la fixation suivante (pour le module visuel) est guidée par ces contraintes locales, indépendamment du parcours effectué jusqu'à présent ou de la page dans sa globalité contrairement au parcours global ou planifié.

Le second parcours que nous appellerons planifié dénote, lui, une activité globale qui opère au niveau de la page, par exemple dans le cas de documents structurés en plusieurs parties. Une vision d'ensemble est alors possible dès la présentation du matériel, indépendamment de la position de la fixation courante. Cette récupération d'informations globale entraîne la mise en place d'un schéma de parcours systématique intégrant tous les éléments à visiter (comme nous l'avons vu dans l'exemple de parcours sur une page Web figure 2.1 page 32). Cette stratégie globale implémente la notion cognitive de *gist*, c'est-à-dire une représentation schématique de la scène dans sa globalité, dès la ou les premières fixations, (Thorpe, Fize, & Marlot, 1996 ; Oliva & Torralba, 2006). Cette représentation cognitive est supposée catégoriser rapidement la scène (par exemple dans le cas d'une scène naturelle, le fait qu'il s'agisse d'une cuisine ou d'un animal) et percevoir l'agencement spatial des objets qui la composent. À partir de cette représentation un schéma de parcours de tous les éléments est généré.

2.4 Processus sémantique

2.4.1 Éléments théoriques

Le type de recherche d'information qui nous intéresse porte sur le sens des mots. Si nous cherchons des informations sur les chevaux, un document comportant les mots *cheval*, *jument* ou *équitation* semble pertinent même s'il ne contient pas exactement le mot-clé demandé.

L'analyse de la sémantique latente (Latent Semantic Analysis, LSA) est un modèle cognitif de la mémoire à long terme qui a été développé pour répondre, entre autres, à ce type de question de façon cognitivement plausible (Landauer & Dumais, 1997). LSA permet de simuler les jugements humains d'associations sémantiques entre groupes de mots. L'application cognitive de cette méthode a été précédée par son utilisation pour la recherche d'information (Deerwester, Dumais, Furnas, Landauer, & Harshmann, 1990). Cette méthode a déjà été utilisée comme application

cognitive, avec le modèle CoLiDeS par exemple.

2.4.2 Analyse de la Sémantique Latente, LSA

Le principe de LSA consiste à définir la signification des mots dans la langue et leurs similarités à partir de l'analyse de larges corpus de texte. Cette signification dépend des différents contextes dans lesquels les mot apparaissent, en se basant sur le fait que des mots qui apparaissent dans des contextes proches sont sémantiquement proches. Une analyse statistique permet de représenter chaque mot par un vecteur dans un espace multidimensionnel de très grande dimension. Tous les mots du corpus sont ainsi représentés par un vecteur et la similarité entre deux mots correspond au cosinus entre les deux vecteurs de ces mots. Les similarités peuvent aussi être calculées entre des groupes de mots, ou entre un mot et un groupe de mots. LSA a un rôle prépondérant dans notre modèle, nous allons donc voir plus en détails son implémentation.

Fondements mathématiques Dans une première étape, LSA prend en entrée un grand corpus de textes, par exemple tous les articles du *Monde* d'une année, et construit une matrice d'occurrences. Dans cette matrice $m \times n$ les lignes représentent les m mots et les colonnes les n documents (nous considérons d'ordinaire comme unité le paragraphe) avec en règle générale $m \gg n$. L'élément f_{ij} de cette matrice A correspond donc au nombre d'occurrences du mot i dans le paragraphe j . Le tableau 2.1 nous montre un exemple simplifié de documents et le tableau 2.2 page ci-contre sa transformation en matrice d'occurrences (Martin & Berry, 2007). Dans cet exemple, les documents sont des titres, dont les mots clés sont en italiques. Les cinq premiers documents parlent de musique et les quatre suivants de cuisine.

TABLE 2.1 – Exemple de documents sur la musique et la cuisine.

Label	Documents
M1	<i>Rock and Roll Music</i> in the 1960's
M2	Differents <i>Drum Rolls</i> , a Demonstration of Techniques
M3	<i>Drum and Bass Composition</i>
M4	A Perspective of <i>Rock Musik</i> in the 90's
M5	<i>Musik</i> and <i>Composition</i> of Popular Bands
B1	How to Make <i>Bread</i> and <i>Rolls</i> , a <i>Demonstration</i>
B2	<i>Ingredients</i> for <i>Crescent Rolls</i>
B3	A <i>Recipe</i> for <i>Sourdough Bread</i>
B4	A Quick <i>Recipe</i> for <i>Pizza Dough</i> using <i>Organic Ingredients</i>

Une seconde étape consiste à normaliser cette matrice afin que les termes les plus informatifs soient privilégiés. Pour chaque élément non-nul de la matrice une

TABLE 2.2 – Matrice d’occurrences obtenue à partir des mots-clés des documents de la table 2.1 page ci-contre.

$$A = \begin{pmatrix} \text{M1} & \text{M2} & \text{M3} & \text{M4} & \text{M5} & \text{B1} & \text{B2} & \text{B3} & \text{B4} \\ \left(\begin{array}{cccccc} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{array} \right) & \begin{array}{l} \textit{Bread} \\ \textit{Composition} \\ \textit{Demonstration} \\ \textit{Dough} \\ \textit{Drum} \\ \textit{Ingredients} \\ \textit{Music} \\ \textit{Recipe} \\ \textit{Rock} \\ \textit{Roll} \end{array} \end{pmatrix}$$

pondération locale (dans chaque document) et une pondération globale (dépendant de tous les documents du corpus) sont appliquées.

$$f'_{ij} = local_{ij} * global_i \quad (2.1)$$

La pondération locale privilégie les termes qui sont les plus fréquents dans le document et sont donc représentatifs de ce document. La pondération globale privilégie les termes qui permettent de différencier les documents entre eux, et sont donc des bons *mots-clés* des documents. Une transformation classiquement utilisée à cette fin est la fonction log-entropie, c’est-à-dire une transformation en log (niveau local) et un calcul d’entropie (niveau global). La fonction locale se définit ainsi :

$$local_{ij} = \log(f_{ij} + 1) \quad (2.2)$$

Cette transformation donne globalement plus de poids aux fréquences élevées tout en diminuant les grandes différences de fréquences.

La fonction globale se définit ainsi :

$$global_i = 1 + \frac{\sum_j p_{ij} \log(p_{ij})}{\log n} \quad (2.3)$$

avec

$$p_{ij} = \frac{f_{ij}}{\sum_j f_{ij}} \quad (2.4)$$

Dans ces équations f_{ij} correspond au nombre d’occurrences du terme i dans le document j , et n correspond au nombre total de documents. Ce calcul de l’entropie donne un poids moindre aux termes qui apparaissent fréquemment dans le corpus entier,

TABLE 2.3 – Matrice d’occurrences normalisée par la fonction log-entropie obtenue à partir de la matrice 2.2 page précédente.

$$A = \begin{pmatrix} \text{M1} & \text{M2} & \text{M3} & \text{M4} & \text{M5} & \text{B1} & \text{B2} & \text{B3} & \text{B4} \\ \left(\begin{array}{cccccc} 0 & 0 & 0 & 0 & 0 & .474 & 0 & .474 & 0 \\ 0 & 0 & .474 & 0 & .474 & 0 & 0 & 0 & 0 \\ 0 & .474 & 0 & 0 & 0 & .474 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .474 & .474 \\ 0 & .474 & .474 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .474 & 0 & .474 \\ .347 & 0 & 0 & .347 & .347 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .474 & .474 \\ .474 & 0 & 0 & .474 & 0 & 0 & 0 & 0 & 0 \\ .256 & .256 & 0 & 0 & 0 & .256 & .256 & 0 & 0 \end{array} \right) & \begin{array}{l} \textit{Bread} \\ \textit{Composition} \\ \textit{Demonstration} \\ \textit{Dough} \\ \textit{Drum} \\ \textit{Ingredients} \\ \textit{Music} \\ \textit{Recipe} \\ \textit{Rock} \\ \textit{Roll} \end{array} \end{pmatrix}$$

		Documents			
		1 2 ...	50 ...	75 ...	100
Corpus 1		aucun X par doc		2 X par doc	
Corpus 2		aucun X par doc		2 X par doc	50 X par doc

FIGURE 2.8 – Illustration de deux corpus ayant le même IDF pour le terme X malgré la différence de distribution du terme.

et ne sont donc pas discriminants. En appliquant cette transformation log-entropie à notre exemple nous obtenons la matrice 2.3.

Cette normalisation peut aussi s’appuyer sur la méthode du TF-IDF (*Term Frequency-Inverse Document Frequency*), qui attribue à chaque mot une valeur dépendant de sa fréquence dans un document (TF, niveau local) et de son importance dans l’ensemble du corpus (IDF niveau global). Mais cette méthode est moins précise car elle ne prend pas complètement en compte la distribution des termes dans le corpus. Prenons par exemple deux corpus avec 100 documents chacun et un terme X. Dans le premier corpus, ce terme est présent 2 fois dans chacun des 50 derniers documents, et dans le second, il est présent 2 fois dans 25 documents et 50 fois dans 25 documents (voir figure 2.8). L’IDF ne capture pas toute l’information, car il prend en compte le fait que le terme X apparaît ou non, sans regarder le nombre d’apparitions. Cette information est donc perdue, et l’IDF des deux corpus pour ce terme sera le même malgré la différence de distribution (Bauer et al., 2005).

La troisième étape de transformation de la matrice est la clé de la méthode. Lors de cette étape cette matrice d’occurrences normalisée est réduite afin de faire émerger les relations entre les termes même si ceux-ci ne sont pas co-occurents. Cette

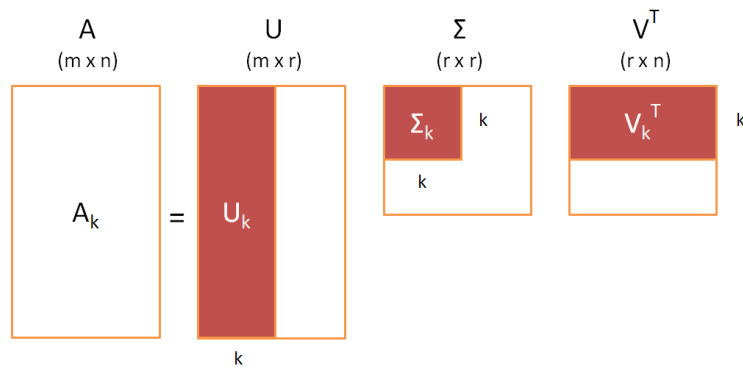


FIGURE 2.9 – Représentation schématique d’une transformation SVD.

réduction peut se faire de plusieurs manières mais celle qui est classiquement utilisée pour LSA est la décomposition en valeurs singulières (*Singular Value Decomposition*, SVD) de la matrice A contenant les termes f_{ij} . Cette matrice A se décompose en un produit de trois matrices :

$$A = U\Sigma V^T \quad (2.5)$$

avec U et V des matrices orthogonales et Σ une matrice diagonale. La figure 2.9 montre une représentation schématique d’une transformation SVD. La matrice U correspond aux termes, et la matrice V aux documents. La matrice Σ contient, elle, les valeurs singulières. En reprenant notre exemple nous obtenons les matrices table 2.4 page 45.

La matrice initiale est réduite en annulant les coefficients de la matrice diagonale, en commençant par le plus petit. Une compression de l’information est ainsi obtenue par la sélection de k dimensions parmi r . Sur notre exemple, et pour une représentation en deux dimensions, seules les deux premières valeurs singulières (les plus grandes) sont conservées (en gras dans la table 2.4 page 45), ce qui nous donne la figure 2.10 page suivante. Sur cette figure, les termes proches sémantiquement le sont spatialement, et ce même s’ils ne co-occurrent pas dans le même document, par exemple *Rock* et *Demonstration*. De la même manière, les similarités entre documents sont observables. Ainsi M1 et M3 sont très proches alors qu’ils n’ont aucun mot en commun. Nous noterons que les mots et les documents se trouvent bien dans le même espace de représentation vectorielle.

Cet exemple simple est significatif de ce que nous pouvons obtenir avec LSA. Le nombre de dimensions (en général $k \approx 300$ dimensions) a été déterminé empiriquement par différents tests, un nombre trop petit conduit à une perte d’information, et un nombre trop grand ne permet pas de faire émerger les liaisons sémantiques entre les mots. Par exemple la figure 2.11 page 46 montre les performances obtenues à un test standard selon le nombre de dimensions utilisées.

Ces dimensions sont non interprétables, même si elles peuvent toutefois être vues comme analogues aux traits sémantiques fréquemment utilisés pour décrire le sens des mots (Landauer et al., 1998).

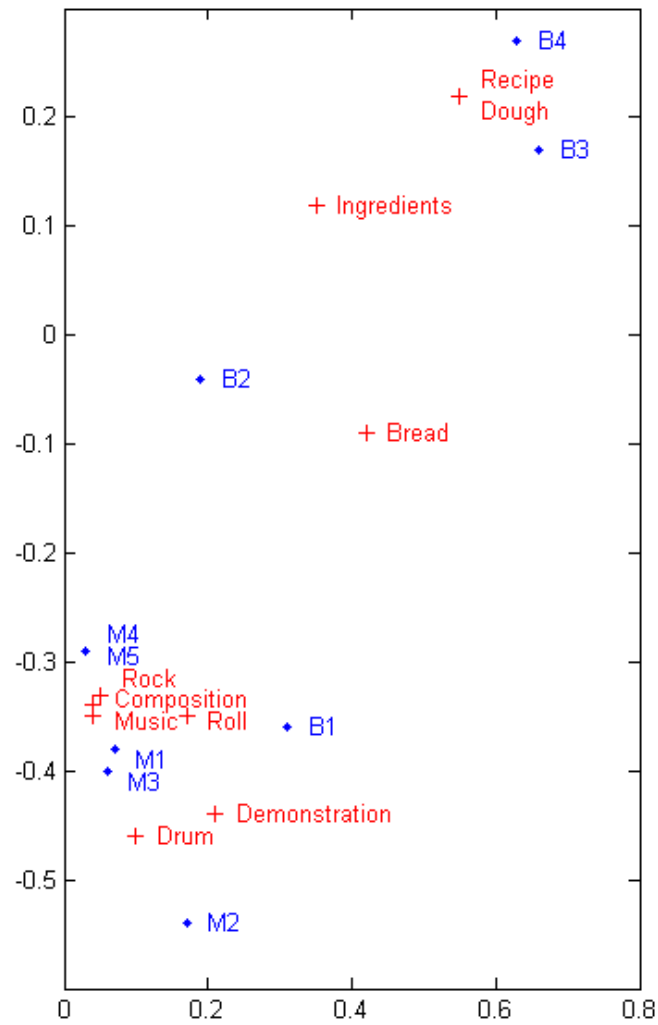


FIGURE 2.10 – Représentation spatiale des données en deux dimensions après la décomposition en valeurs singulières (SVD).

TABLE 2.4 – SVD à partir de la matrice d'occurrences normalisée 2.3 page 42.

Matrice U : Vecteurs des termes

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

Matrice Σ : Valeurs singulières

$$\begin{pmatrix} 1.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .96 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .86 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .76 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .66 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .47 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .27 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .17 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .07 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Matrice V : Vecteurs des documents

$$\begin{pmatrix} .07 & -.38 & .53 & .27 & .08 & .12 & .20 & .50 & .42 \\ .17 & -.54 & -.41 & .00 & .28 & .43 & -.34 & .22 & -.28 \\ .06 & -.40 & -.11 & -.67 & -.12 & .12 & .49 & -.23 & .23 \\ .03 & -.29 & .55 & .19 & -.05 & .22 & -.04 & -.62 & -.37 \\ .03 & -.29 & .27 & -.40 & -.27 & -.55 & -.48 & .21 & -.17 \\ .31 & -.36 & -.36 & .46 & -.15 & -.45 & .00 & -.32 & .31 \\ .19 & -.04 & .06 & -.02 & .65 & -.45 & .41 & .07 & -.40 \\ .66 & .17 & .00 & .06 & -.51 & .12 & .76 & .25 & -.35 \\ .63 & .27 & .18 & -.24 & .35 & .10 & -.35 & -.20 & .37 \end{pmatrix} \begin{array}{l} M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ B1 \\ B2 \\ B3 \\ B4 \end{array}$$

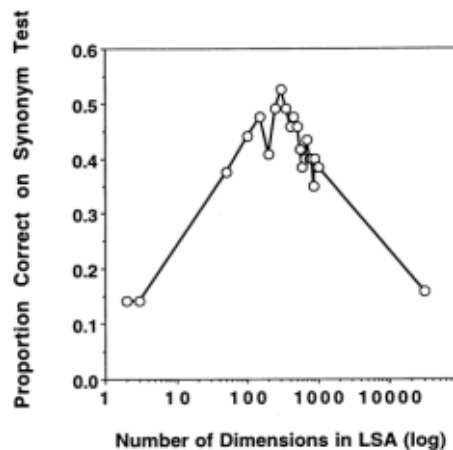


FIGURE 2.11 – Performances au TOEFL en fonction du nombre de dimensions utilisées (Landauer et al., 1998).

Utilisation du modèle Une fois la matrice finale obtenue, les similarités sémantiques entre mots ou groupes de mots peuvent aisément être calculées. Cette représentation vectorielle permet de représenter toute suite de mots par une somme vectorielle des vecteurs des mots qui les composent. Des similarités entre des documents n'appartenant pas au corpus d'origine peuvent être calculées et ce sans avoir à relancer la décomposition en valeurs singulières. Chaque mot ou groupe de mots est défini par un vecteur, et pour calculer la proximité entre ces deux vecteurs il existe plusieurs méthodes. Celle qui est la plus couramment utilisée est le cosinus de leur angle soit une valeur entre -1 et 1, où 1 indique une proximité sémantique maximale.

Une fois toutes ces étapes effectuées, il suffit d'interroger la base et de récupérer les similarités. À partir d'un élément, plusieurs requêtes sont possibles, par exemple la recherche de tous les plus proches voisins sémantiques de cet élément. Ainsi avec comme élément *sport brutal*, et comme corpus français tous les articles parus dans le quotidien *Le Monde* de l'année 1999 (24 millions de mots), la liste suivante des voisins les plus proches est renvoyée (les valeurs de similarités, ainsi que des poids liés à la fréquence, un poids faible indique une forte fréquence, un poids fort une fréquence faible) :

- | | |
|-------------------------------------|--------------------------------------|
| 1. sport sim=0.97 poids=0.42 | 9. sponsors sim=0.67 poids=0.59 |
| 2. sportives sim=0.82 poids=0.55 | 10. athlètes sim=0.66 poids=0.55 |
| 3. sportive sim=0.81 poids=0.52 | 11. hockey sim=0.65 poids=0.71 |
| 4. sportifs sim=0.80 poids=0.51 | 12. marie-george sim=0.61 poids=0.61 |
| 5. sports sim=0.79 poids=0.47 | 13. handball sim=0.61 poids=0.69 |
| 6. compétitions sim=0.75 poids=0.58 | 14. antidopage sim=0.61 poids=0.60 |
| 7. sportif sim=0.69 poids=0.49 | 15. fifa sim=0.60 poids=0.67 |
| 8. tacko sim=0.68 poids=0.94 | ... |

Cette liste est le résultat brut renvoyé par LSA. Il y a donc des mots de fréquence très faible, c'est-à-dire peu présents dans le corpus, comme *tacko*, des dérivations de noms (pluriel, féminin) ou encore des noms propres *Fifa*, *Marie-George*. Chaque mot du corpus a une valeur de similarité avec les mots de la requête, cette liste représente donc les quinze mots les plus proches sur les milliers de mots différents du corpus.

Un des éléments déterminants de cette méthode est le choix du corpus. En effet, selon le corpus, les mots ne vont pas être représentés de la même manière. Dans l'exemple que nous venons de voir, le corpus est constitué d'articles du quotidien *Le Monde*. À l'époque (en 1999) Marie-George Buffet était ministre du sport et c'est la raison pour laquelle son prénom est un proche voisin des mots *sport brutal*. Cette association est clairement liée au corpus. Pour une étude dont les utilisateurs sont des enfants, il faudra utiliser un corpus reflétant les productions écrites auxquelles les enfants ont été exposés. Si le corpus est bien choisi, LSA permet la construction automatique de connaissances sémantiques génériques, simulant ainsi une mémoire sémantique à long terme comparable à la mémoire sémantique humaine. La simulation par LSA de ces connaissances peut être employée quel que soit le domaine dont sont issus les textes pour développer des techniques d'analyse du langage ou enrichir des modèles de traitement de l'information.

Applications du modèle Dans ce mémoire nous utiliserons LSA comme modélisation du sens des mots, mais ce n'est pas la seule application possible. En effet, LSA a aussi été utilisé pour la compréhension de texte (Kintsch, 2000), comme modèle d'acquisition du vocabulaire (Landauer & Dumais, 1997), comme évaluation des connaissances (Lemaire, Mandin, Dessus, & Denhière, 2005) ou encore comme acquisition de connaissances non issues de textes, par exemple sur un jeu de morpion (Lemaire, 1998).

Dans le cas de la modélisation du sens des mots, une des premières études et la plus citée, fut celle réalisée par Landauer et Dumais (1997). Dans leurs travaux les auteurs cherchent à comparer les capacités de leur modèle (LSA) avec celles des humains. Pour ce faire, ils ont pris comme corpus la version numérique de l'encyclopédie *Grolier's Academic American Encyclopedia* qui comptait 4,6 millions de mots (30 473 paragraphes et 60 768 mots différents). Chaque mot fut représenté dans un espace à 300 dimensions. La comparaison entre le modèle et les humains a été faite pour le test de synonymie du TOEFL. Dans cet exercice comprenant 80 items, il s'agit de trouver, pour chaque item donné, celui qui en est le plus proche (le vrai synonyme) parmi les quatre proposés. Par exemple pour le mot *levied* le choix se fait entre :

- | | |
|------------|--------------|
| – imposed | – requested |
| – believed | – correlated |

Les résultats de LSA furent 64,4% (51,5/80) ce qui est très proche des résultats observés chez les étudiants non-anglophones passant ce test afin d'intégrer une école

américaine 64,5% (51,6/80). Ce premier résultat nous montre que ce modèle intègre bien les caractéristiques sémantiques des mots. Pour ce qui est des groupes de mots, il a fallu aussi vérifier que l'addition des vecteurs des différents mots d'un groupe avait un sens, et que le vecteur ainsi généré correspondait bien à la signification du groupe de mots. En comparant la capacité du modèle à discriminer des textes selon leur difficulté et celle des humains, Rehder et al. (1998) ont validé la représentation sémantique des groupe de mots telle que produite par LSA.

Limites de LSA

Nous noterons deux principales limites à cette méthode. Tout d'abord le principe selon lequel les textes sont traités comme des *sacs de mots*. Toute syntaxe est perdue, alors que le sens des propositions peut dépendre de leur syntaxe. Par exemple, le vecteur LSA associé à *réforme de l'enseignement* est le même que pour *enseignement de la réforme*. Cette problématique se pose surtout pour des propositions courtes car sur un texte plus long un tel renversement des rôles et donc du sens ne se fait pas aussi facilement. Dans le cas de documents très courts (deux ou trois mots) une validation supplémentaire des similarités sémantiques par des participants peut-être nécessaire.

La seconde limitation que nous évoquerons ici concerne le fait que chez l'homme les connaissances en mémoire à long terme ne sont pas issues que de textes. Nous créons ces connaissances à partir de nos expériences perceptuelles (visuelles, tactiles, auditives) et de nos actions. Glenberg et Robertson (2000) ont fait une comparaison entre les réponses de LSA et de participants sur des situations prenant en compte cet aspect vécu de l'apprentissage des concepts. Si la situation est la suivante, quelle est la meilleure réponse ? :

Brad était assis dans son bureau quand un intrus l'a menacé avec un couteau

- a. *Brad a attrapé son ordinateur portable afin de se protéger.*
- b. *Brad a attrapé sa serviette en papier afin de se protéger.*
- c. *Brad a attrapé son fusil de chasse afin de se protéger.*

Pour LSA la 3ème réponse est la meilleure, mais il ne fait pas de différences entre les deux premières. Ainsi il n'a pas été capable d'extraire le fait qu'un ordinateur est rigide et peut servir de bouclier. Dans cet article, plusieurs expérimentations montrent cette limitation de LSA.

Ces limitations se retrouvent dans les autres modèles de représentation sémantique des mots que nous allons évoquer.

Autres modèles sémantiques

LSA n'est pas le seul modèle de représentation sémantique des mots, d'autres modèles existent, comme WordNet et PMI-IR.

WordNet WordNet (Miller, 1995) est une taxinomie de l'anglais regroupant un très grand nombre de mots ainsi que les liens qu'ils entretiennent entre eux. Les mots

sont organisés en ensembles de synonymes appelés *synsets*. WordNet est structuré par des relations entre synsets et entre mots. Ces relations sont la synonymie, l'hyponymie (englobant), l'hyponymie (inverse de l'hyponymie), la méronymie (partie de) et l'holonymie (inverse de la méronymie). Les mots dans WordNet peuvent donc être représentés sous forme de graphe et la similarité entre mots dépend de la distance entre eux dans ce graphe (nombre d'arcs par exemple). Cette représentation est cependant codée à la main et en anglais ce qui limite son utilisation.

PMI-IR Dans PMI-IR (*Pointwise Mutual Information - Information Retrieval*) l'idée de base est que le sens d'un mot est caractérisé par ce qui l'entoure. L'approche statistique choisie utilise l'information mutuelle entre les mots et les groupes de mots et est basée sur les co-occurrences de mots. Le corpus utilisé est le web (plusieurs millions de pages) puisque la mesure de co-occurrence utilise un moteur de recherche. Cette méthode a, elle aussi, été utilisée sur des tests de synonymie du TOEFL. Pour un mot *problème* donné, par exemple *levied* le but est de choisir, parmi les quatre mots proposés, celui qui est le plus proche sémantiquement du *problème*. L'information mutuelle est définie comme suit :

$$score(choix_i) = \frac{p(\text{problème}\&\text{choix}_i)}{p(\text{choix}_i)} \quad (2.6)$$

$p(\text{problème}\&\text{choix}_i)$ est la probabilité que *problème* et *choix_i* co-occurrent. La définition de la co-occurrence est affinée pour ne prendre que les mots co-occurants qui sont proches spatialement l'un de l'autre (moins de dix mots entre eux). Une autre modification permet de favoriser les synonymes face aux antonymes, et enfin le contexte peut-être pris en compte.

Cette méthode a été comparée à LSA par Turney (2001) et a montré un taux de bonnes réponses au test des synonymes du TOEFL 10% supérieur à LSA (73,75%, 59/80). Mais cette performance est aussi 10% supérieure aux données expérimentales, et ne simule donc pas correctement ces données.

Comparaison Kaur et Hornof (2005) ont comparé ces trois modèles (LSA, WordNet et PMI-IR) avec les performances des humains. Dans la tâche étudiée les utilisateurs devaient choisir les liens à cliquer lors d'une navigation, étant donné une page Web et un but. Les liens et les buts pouvant être composés de plusieurs mots, deux versions de PMI-IR ont été testés. La première, PMI-w, renvoie comme valeur la meilleure similarité parmi toutes les similarités entre chaque mot du but et du lien. La seconde, PMI-p, renvoie comme valeur la similarité entre les groupes de mots. Le problème avec PMI-p, est que s'il s'agit de phrases entières elles peuvent ne pas être dans le corpus, ou sous-représentées. En effet, dans PMI-IR il n'y a pas de représentation vectorielle et un document qui n'apparaît pas dans le corpus initial n'a pas de valeur. La comparaison entre les quatre modèles a montré une meilleure performance pour PMI-w.

La supériorité de PMI-w est cependant discutable, car cette mesure ne prend en compte que la meilleure similarité entre tous les mots du but et du lien, il ne s'agit donc au final que de la similarité entre deux mots. Cette mesure n'est donc pas applicable entre des documents plus longs, car cela sous-entendrait que chaque document se résume à un mot. De plus, les performances obtenues sont meilleures que celles obtenues par les utilisateurs, il ne s'agit donc pas d'une bonne simulation du comportement observé.

2.4.3 Piste de modélisation

Pour la modélisation du processus sémantique nous séparerons en deux catégories les traitements effectués par le modèle, tout en restant à un niveau local. Nous verrons dans la section modélisation le traitement et le parcours entre éléments lorsque ceux-ci sont des mots isolés. Dans les pages Web, nous retrouvons un principe de cohérence générale dans l'organisation de la page, qui induit une proximité spatiale des éléments proches sémantiquement, nous aurons cette même idée dans le cas de mots isolés. Nous verrons ensuite le traitement de paragraphes, assimilé à un mécanisme de lecture. Nous nous inspirerons de l'idée d'*information scent* et du fait que lorsque l'information dans une page n'est plus assez importante il est temps de changer de page. Nous appliquerons cette idée au niveau des paragraphes.

Dans les deux cas nous utiliseront LSA comme modèle de la mémoire sémantique à long terme. LSA reste en effet le modèle qui a été comparé dans de nombreux cas aux performances humaines avec de bons résultats. Notre objectif est d'être le plus proche possible des humains et non le plus performant, LSA avec une approche par corpus est cognitivement plus plausible. De plus, une fois le corpus choisi il permet la comparaison entre des phrases ou des paragraphes (même si ceux-ci ne font pas initialement partie du corpus) et pas simplement entre des mots.

2.5 Processus mnésique

2.5.1 Éléments théoriques

La mémoire est un processus fondamental en sciences cognitives et notamment en recherche d'information. Il s'agit de la faculté de traiter les informations perçues, afin de pouvoir les réutiliser par la suite. La notion de mémoire regroupe des processus différents : l'encodage, le stockage et la récupération des informations.

Un découpage classique sépare la mémoire à long terme et la mémoire à court terme (Atkinson & Shiffrin, 1968). La troisième composante identifiée dans cette théorie est le registre sensoriel dans lequel une grande quantité d'information (notamment visuelle) peut être stockée pendant un temps très court (quelques millisecondes).

La mémoire à long terme se divise elle-même selon deux dimensions. Une première distinction est faite entre la mémoire procédurale et déclarative, c'est à dire

entre les savoir-faire et les savoirs. La seconde distingue la mémoire épisodique de la mémoire sémantique, la première concerne l'histoire de l'individu, ce qu'il a vécu, les dates, les événements, et la seconde les connaissances générales, concepts, définitions, qui se rapporte, entre autre, au langage. Ces distinctions se fondent sur des expérimentations en laboratoire et des études de patients cérébro-lésés.

En mémoire à court terme les éléments ne sont stockés qu'un temps très court (entre 15 et 30 secondes) et cette mémoire est un espace intermédiaire entre la mémoire à long terme et l'environnement. Le concept de mémoire de travail permet de mieux rendre compte de cette notion d'espace dans lequel les informations sont effectivement utilisées. La mémoire de travail se divise elle-même en trois composantes d'après le modèle de [Baddeley et Hitch \(1974\)](#) : l'administrateur central, la boucle phonologique et le calepin visuo-spatial. L'administrateur central contrôle et coordonne les deux autres systèmes. La boucle phonologique sert à retenir et traiter les informations sous forme verbale. Le calepin visuo-spatial traite les informations sous forme visuelle.

À partir de ces définitions succinctes, nous avons construit notre modèle en simulant deux types de mémoire : la mémoire sémantique et la mémoire de travail.

La mémoire sémantique correspond aux connaissances sémantiques stockées en mémoire et modélisées avec LSA dans le processus sémantique. La mémoire de travail est quant à elle utilisée lors de la recherche d'information afin de mettre en place une stratégie de recherche. Cette mémoire permet en effet à l'homme d'être efficace dans sa recherche en ne revenant pas sur ce qui a déjà été traité si cela n'est pas nécessaire. Nous verrons quels sont les différents points de vue sur la capacité de cette mémoire en recherche d'information, ainsi que deux modèles cognitifs que nous retrouvons dans la littérature. Le mécanisme d'inhibition de retour, combiné à un mécanisme d'oubli et le *Variable Memory Model*.

La capacité de la mémoire de travail en recherche d'information

Quelle est la capacité de la mémoire lors d'une recherche visuelle ? Deux types d'informations peuvent être stockées en mémoire, *où* et *quoi*. Lorsque nous considérons les deux conjointement, il semble que la mémoire soit limitée à trois ou quatre paires d'éléments (localisation + identité) ([Horowitz & Wolfe, 2001](#)). Si nous nous intéressons essentiellement à la mémorisation du trajet et donc aux localisations visitées, celle-ci apparaît être d'une grande capacité (plus de douze places retenues), mais avec une faible résolution (l'identité de l'objet n'est pas précise) ([Dickinson & Zelinsky, 2007](#)). D'autres auteurs parlent même d'une mémoire allant jusqu'à 20-25 distracteurs rejetés ([Takeda, 2004](#)).

À partir de ce constat nous avons tout d'abord choisi d'implémenter une mémorisation très précise des derniers items visités, qui se dégrade ensuite avec le temps. Pour ce faire, nous nous sommes intéressés au mécanisme d'inhibition de retour, classiquement modélisé en vision, associé à un mécanisme d'oubli.

Inhibition de retour (IOR)

Lors d'une recherche d'information, nous devons déplacer notre attention sur la scène, créant ainsi une séquence de fixations. Dans les différents modèles qui tentent de reproduire les processus qui influencent l'emplacement de ces fixations, qu'ils soient principalement ascendants (Itti & Koch, 2000 ; Rosenholtz, Li, Mansfield, & Jin, 2005) ou intégrant des processus descendants (Navalpakkam & Itti, 2005), l'attention se déplace entre les régions en partant de la plus attractive. Afin de ne pas rester sur la région la plus saillante le mécanisme choisi est l'inhibition de retour. Ce mécanisme diminue le poids des régions déjà fixées.

L'IOR est une tendance observée expérimentalement contre la réorientation de l'attention vers des localisations déjà visitées. Lors de présentation de cible à l'emplacement de la fixation précédente, les temps de réaction ou les durées latentes pour initier la saccade quittant cet emplacement sont plus longs. Ce phénomène a été signalé pour la première fois par Posner et Cohen (1984), et a été répliqué en recherche visuelle (Klein & MacInnes, 1999). Il a été proposé que cette fonction d'inhibition facilite la recherche visuelle. Cette hypothèse est étayée par le fait qu'au cours de recherche visuelle les saccades oculaires ont tendance à être dirigées loin de l'emplacement de la fixation précédente et que la durée des fixations précédant un retour sur cet emplacement est plus longue que si la saccade s'éloigne de cet emplacement (Klein & MacInnes, 1999).

Ce mécanisme a néanmoins ses limites. Hooge, Over, van Wezel, et Frens (2005) ont examiné les mouvements de l'œil au cours de l'exploration de scènes visuelles et ont découvert que le nombre de refixations sur l'emplacement de la fixation précédente est plus important que ne le prédit le hasard. De même, une analyse des tendances systématiques lors de perception de scène (saccades et fixations) a montré cette facilitation de retour (Tatler & Vincent, 2008). Une forte fréquence de ces refixations précisément à l'endroit de la fixation précédente a aussi été démontrée sur différentes tâches effectuées par des singes (Motter & Belky, 1998). L'exactitude de ces refixations suggère que le retour de l'attention à l'emplacement précédemment fixé peut être facilité plutôt qu'inhibé. La facilitation du retour pourrait être approprié pour des scènes complexes ou des tâches complexes (Henderson & Smith, 2009).

Les causes de ces refixations peuvent être multiples, la refixation peut survenir en raison de l'insuffisance de traitement de l'élément fixé, d'un problème de représentation en mémoire ou tout simplement parce que la région est très informative. L'information à l'emplacement précédemment fixé peut ne pas avoir été complètement encodée ou la similarité avec un objectif de recherche peut nécessiter un retour, ou encore la mémorisation de l'objet fixé peut s'être détériorée. En prenant en compte ces éléments nous avons examiné une autre approche de modélisation de la mémoire impliquée dans des tâches de recherche visuelle. Le *Variable Memory Model*, décrit par Arani et al. (1984) et repris par Horowitz (2006) semble être un bon candidat.

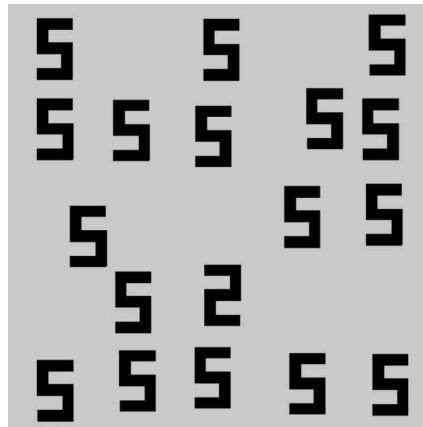


FIGURE 2.12 – Exemple d’image utilisée dans les expérimentations d’Horowitz (2006). Les participants doivent chercher le 2 parmi les 5.

Variable Memory Model (VMM)

La question que se sont posés les auteurs est la suivante : dans quelle mesure le système visuel se rappelle des objets qui ont déjà été examinés au cours d’une tâche ?

Les hypothèses de ce modèle sont les suivantes :

- sélection d’un item à la fois (recherche sérielle) ;
- processus attentionnel conscient (*overt attention*).

Le principe du VMM est d’utiliser des probabilités d’encodage et de récupération des informations en mémoire. Un item visité sera encodé avec une probabilité θ et un item encodé sera récupéré avec une probabilité Φ , en fonction du nombre de pas entre son encodage et sa récupération. Dans son article, l’auteur utilise une comparaison entre des données expérimentales et celles du modèle pour définir ces probabilités sur une tâche de recherche visuelle (Horowitz, 2006). Les participants doivent par exemple chercher un 2 parmi des 5 (voir figure 2.12). Les meilleures valeurs qu’il obtient pour simuler les observations avec son modèle sont : $\theta = 0,82$ et $\Phi = 0,86$.

Grâce à ce modèle, des refixations sont possibles, ce qui est en accord avec les données empiriques.

2.6 Modélisation

Après avoir vu les différents éléments théoriques et quelques pistes de modélisation, nous allons maintenant présenter dans cette section l’implémentation des trois processus.

2.6.1 Processus visuel

Modélisation du parcours réactif

Le premier mécanisme qui influence le poids des éléments dans la carte visuelle est la saillance visuelle des stimuli, représentée par une carte de saillance qui pourrait

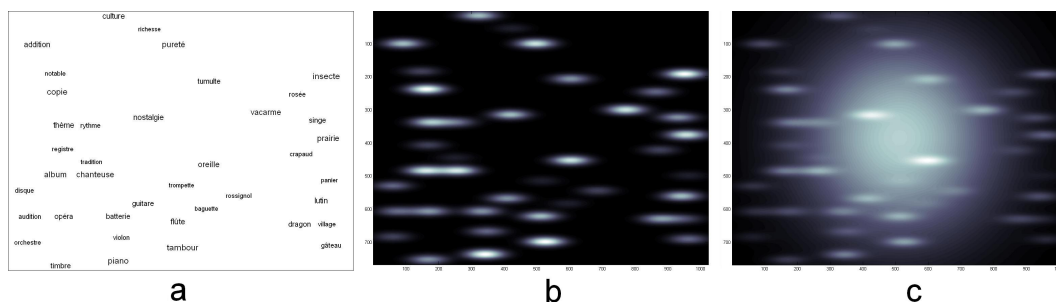


FIGURE 2.13 – (a) Page originale, les mots ont des tailles plus ou moins grandes (de 13 à 19) réparties aléatoirement ; (b) Carte de saillance visuelle statique associée à la page, la saillance des mots dépend de leur taille, avec une acuité constante ; (c) Carte de saillance visuelle modulée par un filtre rétinien, ici au centre de la page.

être comparée à celles générées par [Itti et Koch \(2000\)](#), présenté dans le chapitre précédent. Pour une page, nous obtenons donc une carte, avec un poids pour chaque élément, représentée sous forme de carte de chaleur (exemple figure 2.13 (a) et (b)). Dans les expériences que nous présenterons par la suite, et que nous comparerons au modèle, la saillance visuelle des éléments sera très simple, et la carte de saillance visuelle ne sera donc pas aussi élaborée que dans le modèle cité. Elle se limitera à une pondération des éléments saillants ou considérés comme saillants a priori. Par exemple si un élément est en couleur il sera plus saillant qu’un élément sans couleur.

La seconde propriété visuelle utilisée dans ce modèle est l’acuité visuelle, que nous avons vu précédemment (section 2.3.1 page 37). Cette propriété est très importante car elle permet de prendre en compte la localisation courante, ce qui n’est pas le cas de la plupart des modèles de guidage de l’attention.

Pour modéliser cette propriété nous nous sommes basés sur les travaux de [Geisler et Perry \(2002\)](#). Un filtre rétinien est simulé, correspondant à l’acuité visuelle humaine, et suivant la courbe figure 2.14 page suivante. Cette courbe est une modélisation de la répartition des cônes sur la rétine humaine (voir figure 2.6 page 38). Ainsi la zone fovéale autour du point de fixation est très précise alors que les zones parafovéales et périphériques ont une acuité beaucoup plus faible. Ce filtre se déplace dynamiquement autour du point de fixation et module la carte de saillance visuelle précédemment décrite. Par exemple sur l’image (c) figure 2.13 ce filtre est situé au milieu de l’image.

Ainsi un élément saillant trop loin de l’élément courant aura un poids visuel amoindri, et au contraire un élément proche aura un poids renforcé même s’il n’est pas particulièrement saillant. La carte obtenue après application de ce filtre est dépendante du point de fixation courant, elle est donc générée dynamiquement tout au long du parcours.

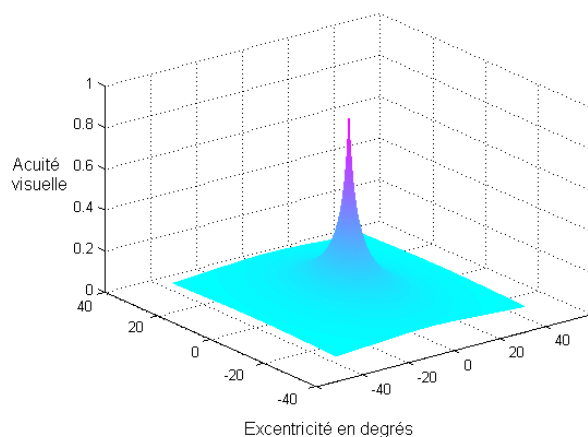


FIGURE 2.14 – Courbe d’acuité visuelle selon l’excentricité.

Modélisation du parcours planifié

Le parcours que nous appellerons planifié, ou global, se construit à partir de la page dans sa globalité avant la première fixation (notion de *gist*). Ce parcours peut se construire à partir de tout type de matériel. Nous l’appliquerons ici dans le cas de documents structurés en paragraphes. Comme nous le verrons lors de l’observation des données dans les chapitres suivants les humains ont tendance à avoir des stratégies de recherche dépendant de la structure des documents et à faire un parcours systématique de tous les éléments. Le modèle intègre donc ce mécanisme en ayant une vision globale de la page dès le début de la recherche et en en prévoyant le parcours adéquat. Ce mécanisme ne se représente pas sous forme d’une carte en attribuant à chaque élément un poids, mais plutôt sous forme d’une liste d’éléments à parcourir dans un certain ordre. À chaque nouvel élément traité le modèle génère de nouveau le meilleur chemin à suivre pour le reste des éléments.

Dans notre modèle, ce principe est modélisé en construisant à l’avance un schéma de parcours, qui sera adapté dynamiquement après chaque visite. Nous avons modélisé ce problème à l’aide d’une analogie avec le problème du voyageur de commerce, largement étudié dans la littérature. Il consiste à relier un ensemble de points en les visitant tous une seule fois et en minimisant la longueur du parcours.

Algorithme basé sur celui du voyageur de commerce L’objectif est de produire un chemin visitant chaque élément, avec un critère de coût dépendant ici de contraintes cognitives. Plusieurs travaux ont essayé de caractériser l’algorithme qui décrirait le mieux la méthode humaine de résolution du problème du voyageur de commerce. L’algorithme d’insertion du nœud le moins coûteux dans l’enveloppe convexe semble être un bon candidat (MacGregor, Ormerod, & Chronicle, 2000). Il s’agit de construire le parcours élément par élément, en y insérant au fur et à mesure l’élément le plus proche du point de départ (et non encore inséré) au meilleur endroit, c’est-à-dire là où il entraîne la plus faible augmentation du coût. Voyons par

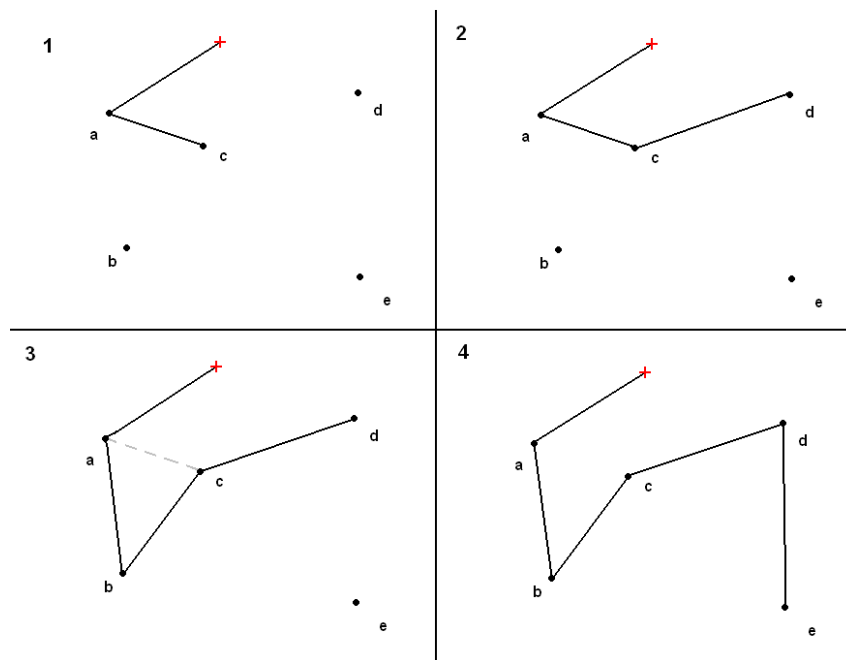


FIGURE 2.15 – Simulation de la construction d'un parcours planifié sur une page à cinq éléments. 1 : les éléments a et b sont insérés. 2 : l'élément c est inséré entre a et b pour minimiser les coûts. 3 et 4 : les éléments d et e sont insérés.

exemple la simulation d'un parcours planifié sur une page à cinq éléments $\{a,b,c,d,e\}$ (voir figure 2.15). Si nous insérons ces éléments en commençant par le plus proche du point de départ (la croix rouge), le parcours sera "a", puis "a-c" (1 : en haut à gauche de la figure), puis d est inséré car c'est l'élément suivant le plus proche (2 : en haut à droite de la figure). Puis en insérant b, nous trouverons que la meilleure place est entre a et c, ce qui donnera le trajet "a-b-c-d" (3 : en bas à gauche de la figure). Pour e l'insertion se fera aussi en limitant les coûts et nous obtenons finalement le parcours planifié "a-b-c-d-e".

Lorsque le parcours est construit, le modèle se dirige vers le premier élément du chemin. Quand il a terminé le traitement de cet élément, il reconstruit un nouveau chemin intégrant tous les éléments non visités, à partir de l'élément courant. Le coût de l'insertion d'un élément dépend de la distance entre les éléments mais aussi de préférences heuristiques sur les directions.

Chez les humains dont la langue maternelle s'écrit et se lit de gauche à droite et de haut en bas, il existe une tendance à parcourir les documents selon ces directions, comme nous l'avons aussi observé dans nos expériences avec des participants de langue maternelle française. Nous avons repris ces paramètres dans la modélisation en pondérant les directions.

Le modèle a deux paramètres α_H et α_G qui correspondent respectivement au poids relatif d'un déplacement vers le haut, et au poids relatif d'un déplacement vers la gauche.

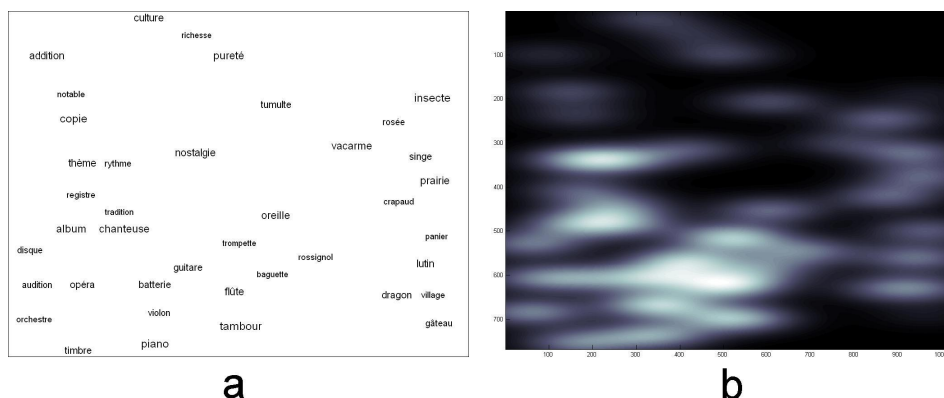


FIGURE 2.16 – (a) Page contenant des mots organisés : les mots proches sémantiquement le sont aussi spatialement ; (b) Carte de saillance sémantique statique associée à la page, la saillance des mots dépend de leur similarité sémantique avec la question, qui est ici *Quel est l'instrument de musique le plus petit ?*

Si α_H vaut 2, une transition vers le haut est, pour la même distance, deux fois plus coûteuse qu'une transition vers le bas. De même, si α_G vaut 1, une transition vers la gauche est, pour la même distance, aussi coûteuse que vers la droite. Cet algorithme peut être appliqué à différentes configurations comme nous le verrons dans les chapitres suivants.

Nous prenons donc en compte à un niveau global la structure du document, et à un niveau local la saillance visuelle des éléments ainsi que l'acuité visuelle dépendant de l'excentricité des éléments par rapport à l'élément courant.

2.6.2 Processus sémantique

Application à un environnement simple

Nous allons commencer par considérer que les éléments sont des mots et partir du principe que plus un élément est proche sémantiquement du but de la recherche plus les éléments qui l'entourent le sont aussi. Ce principe découle du fait que dans le monde qui nous entoure les éléments sont rangés et ordonnés de manière cohérente, nous ne trouvons pas de dentifrice au milieu des fruits et légumes, ni d'information sur des hôtels dans un forum de programmation. Ce principe se retrouve aussi au sein d'une même page Web. La figure 2.16 illustre ce phénomène. Sur cette figure (a) quarante mots sont organisés sémantiquement. Pour une question donnée les éléments proches de cette question sont proches les uns des autres (b).

L'idée est donc de modéliser une tendance à quitter la zone courante si l'élément fixé est sémantiquement éloigné du but et, au contraire, à y rester si l'élément fixé est proche du but. Pour calculer les similarités sémantiques nous utilisons LSA, qui est comme nous l'avons vu précédemment un bon estimateur des jugements humains d'associations sémantiques.

À partir de la fixation courante, la carte sémantique est modulée en appliquant

un filtre gaussien dépendant de la valeur du mot fixé. Ainsi si le mot est proche de la question son entourage proche va être avantagé, incitant le modèle à rester dans cette zone. Si le mot est éloigné, son entourage va être pénalisé, incitant le modèle à quitter cette zone. La valeur seuil entre l'avantage ou la pénalité se trouve lorsque nous considérons que le mot n'est plus relié sémantiquement à la question. Ce seuil est estimé empiriquement et fixé à 0,2. En effet, à partir de cette valeur de similarité nous considérons généralement qu'il existe un lien sémantique suffisant entre les éléments (Katsanos, Tselios, & Avouris, 2009).

Le filtre utilisé a la forme d'une gaussienne, mais sa hauteur (cloche vers le haut ou vers le bas) dépend de la similarité entre le mot fixé et le but. Elle est maximale lorsque le cosinus LSA est à 1 (similarité maximale), nulle lorsque le cosinus LSA vaut 0,2 et minimale lorsque le cosinus vaut 0. La fonction la plus simple répondant à ces critères est illustrée figure 2.17 page ci-contre, par exemple pour une similarité de 0,45 la hauteur sera égale à 0,25.

Une fois la hauteur déterminée le filtre est appliqué à tous les mots de la page. Ceux qui sont proches de la fixation courante sont fortement impactés et les mots plus éloignés le sont moins. Le poids sémantique des mots devant se situer entre 0 et 1, un ajout de 0,5 est donc effectué pour n'avoir que des valeurs positives. Ainsi nous pouvons obtenir, selon la similarité LSA du mot fixé avec le but, des gaussiennes de différentes formes (voir quelques exemples figure 2.18 page suivante), par exemple pour une similarité de 0,45 la cloche est vers le haut et a un maximum à 0,75 (0,25+0,5).

L'équation de la gaussienne, une fois la hauteur déterminée est la suivante :

$$poidsSem = 0,5 + (hauteur \times e^{-(2 \times distAFixCourante)^2}) \quad (2.7)$$

Application à un environnement complexe

Décrivons maintenant le modèle lorsque les éléments ne sont plus des mots mais des suites de mots (paragraphes). Dans ce cas, notre modèle effectue un traitement sémantique assimilé à de la lecture mot à mot afin de décider si celui-ci est en lien ou non avec le but de la recherche. Grâce à la représentation vectorielle de LSA nous pouvons appliquer les mêmes traitements à des mots ou des paragraphes. Il s'agit donc de modéliser la compréhension du texte ainsi que de calculer la similarité entre le texte traité et le thème de la recherche.

Nous avons de nouveau utilisé LSA comme modèle de la mémoire à long terme pour nous permettre d'évaluer la pertinence de la séquence de mots traités par rapport au but de la recherche d'information, de manière dynamique. Chaque suite de mots est représentée par un nouveau vecteur qui peut être comparé au vecteur « thème » de la recherche par une simple mesure de cosinus. Par exemple si le modèle a traité les mots « les via ferrata et les sports eau vive connaissent un succès » et que le thème de la recherche est *tourisme en montagne* la similarité LSA est 0,38.

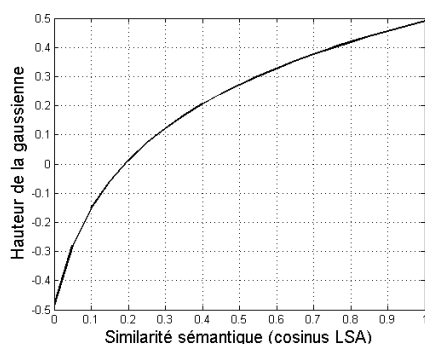


FIGURE 2.17 – Courbe définissant selon le cosinus du mot fixé la hauteur de la gaussienne modulant la carte sémantique, à 0,2 la gaussienne sera nulle, en-dessous elle sera vers le bas, et au-dessus elle sera vers le haut.

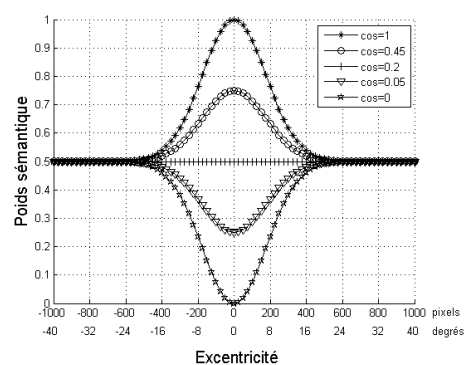


FIGURE 2.18 – Exemples de gaussiennes pour quatre valeur de cosinus différents. Cette gaussienne module la carte sémantique dans un sens ou l'autre selon la valeur du cosinus de l'élément courant.

Comme nous le verrons par la suite avec les différentes expérimentations, certains mots peuvent avoir une fonction sémantique particulière, leur importance et leur prise en compte seront donc différentes.

Le processus par lequel l'homme recueille des informations dans un texte est simulé ici en considérant les mots un à un. Dans ce type de recherche, certains utilisateurs ne lisent pas vraiment les paragraphes linéairement et préfèrent chercher des mots ici et là, mais la plupart commencent un paragraphe sur les premiers mots de la première ligne. Nous utiliserons donc ce comportement simplifié dans notre modèle. De plus, ce qui est très important à simuler est le fait que les utilisateurs ne lisent pas systématiquement les paragraphes dans leur intégralité. S'ils ont acquis assez d'information, ils décident de quitter le paragraphe. Il s'agit donc de modéliser la décision appropriée. Il y a trois conditions qui permettent de modéliser au plus simple la décision d'arrêter ou de continuer la lecture d'un paragraphe :

- Si les mots traités jusqu'à présent ne sont pas liés à ce thème (en utilisant les mesures de similarité sémantique de LSA), le bloc est abandonné avec une valeur égale à zéro, signifiant sa non pertinence pour les revisites ;
- Si les mots traités jusqu'à présent sont sémantiquement très proches du thème, le bloc est abandonné, mais il est stocké en mémoire comme une solution possible avec une valeur correspondant à sa similarité sémantique avec le thème ;
- Si les mots traités jusqu'à présent ne sont ni très proches, ni très éloignés du thème, la lecture continue et le mot suivant est traité.

Le modèle continue sa lecture jusqu'à ce qu'il puisse prendre une décision ou jusqu'à la fin du paragraphe. Deux seuils de similarité sémantique ont été définis à cet effet pour définir ce que sont une grande et une faible proximité sémantique. Ces seuils varient selon le nombre de mots déjà traités dans le bloc : plus le modèle a vu

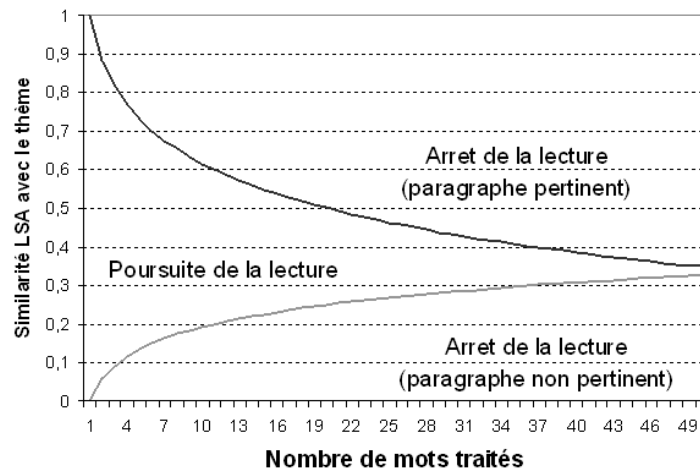


FIGURE 2.19 – Seuils de similarité en fonction du nombre de mots traités

de mots, plus il peut être capable de prendre sa décision. Par exemple, si seulement deux mots ont été traités, les seuils sont très restrictifs car la décision d’abandonner le bloc ne doit être prise que si la similarité entre ces deux mots et le thème est extrêmement élevée ou extrêmement faible. Par contre si 25 mots ont été traités, il y en a assez pour avoir une meilleure confiance dans la mesure sémantique. Les seuils sont définis de la manière suivante (P est un bloc, S est la séquence de mots traités jusque là et T est le thème de recherche) :

$$Abandon(P) \text{ si } sim(S, T) < \frac{\log(nbMots(S))}{12} \quad (2.8)$$

$$Abandon(P) \text{ si } sim(S, T) > \frac{1 - \log(nbMots(S))}{6} \quad (2.9)$$

Ces fonctions ont été définies expérimentalement et ont donné un comportement approprié. La figure 2.19 présente ces seuils en fonction du nombre de mots traités. Si la similarité avec le thème se situe entre les deux lignes, le modèle continue sa lecture et vérifie le mot suivant. Dans les autres cas, le bloc est abandonné et la simulation reprend au niveau des blocs. L’abandon est la plupart du temps lié à la non pertinence du bloc.

Ce traitement sémantique aura des conséquences sur la suite du trajet. Un élément jugé intéressant sera plus à même d’être revisité qu’un élément jugé inintéressant, c’est ce que nous verrons dans la modélisation du mécanisme mnésique.

2.6.3 Processus mnésique

Au niveau de la mémoire le premier principe est d’aller sur les éléments qui n’ont pas été traités afin d’avancer dans la recherche.

Application à un environnement simple

Comme nous l'avons mentionné précédemment les humains effectuent de nombreuses saccades régressives revenant sur la localisation fixée précédemment. Dans notre implémentation du modèle nous distinguons les éléments visités et ceux traités, ainsi un élément visité mais non traité ne sera donc pas encodé en mémoire et le modèle aura la possibilité d'y retourner. C'est le principe du *Variable Memory Model*.

Chaque mot a un poids qui correspond à son attractivité. Le poids initial de tous les éléments est égal à 1, aucun mot n'est encodé en mémoire. Plus le poids est grand, plus cet élément est attractif, au départ tous les éléments ont donc une attraction maximale.

En quittant un élément si l'encodage a été effectué (dépendant de la probabilité θ) son poids diminue fortement. Il commencera ensuite à remonter progressivement. Dans le cas où il n'a pas été encodé, son poids est toujours égal à 1. À chaque pas tous les éléments qui ont déjà été vus et encodés voient leur poids remonter selon la courbe dépendant de Φ . Un élément qui a été vu mais non encodé le sera lors du second passage (s'il a lieu). Le mécanisme d'oubli décrit dans le *Variable Memory Model* correspond à la première courbe figure 2.20 page suivante. Sur cette courbe une valeur proche de 0 signifie que le mot vient d'être vu (et donc que le modèle n'a pas tendance à y retourner) une valeur à 1 signifie au contraire que ce mot n'a jamais été vu, et attire donc le modèle. Ce mécanisme est formalisé par l'équation suivante :

$$poidsMem_{i,n} = 1 - \Phi^n \quad (2.10)$$

avec i représentant l'élément, n son rang dans le trajet et Φ la probabilité du modèle de mémoire.

Dans une première version du modèle, nous avons implémenté le classique mécanisme d'inhibition de retour, couramment utilisé dans les modèle de guidage de l'attention, associé à un mécanisme d'oubli (Chanceaux, Guérin-Dugué, Lemaire, & Baccino, 2008). Ce mécanisme, qui s'est révélé une simplification du mécanisme précédent était modélisé par l'équation suivante et correspond à la seconde courbe figure 2.20 page suivante :

$$poidsMem_{i,t} = poidsMem_{i,t-1} + \frac{(1 - poidsMem_{i,t-1})}{\delta} \quad (2.11)$$

δ a été expérimentalement fixé à 6 dans une première étude (Chanceaux et al., 2008), t est l'instant courant, le poids du mot i dépend de son poids au temps $t - 1$.

Cette courbe modélise le fait qu'un élément est inhibé lorsqu'il vient d'être fixé (poids à 0), puis son attractivité remonte au fur et à mesure, jusqu'à être complètement oublié au bout d'une vingtaine d'étapes.

Afin d'être plus proche du comportement observé, nous avons aussi pris en compte le poids sémantique des éléments lors de la phase d'encodage. En effet, si un élément

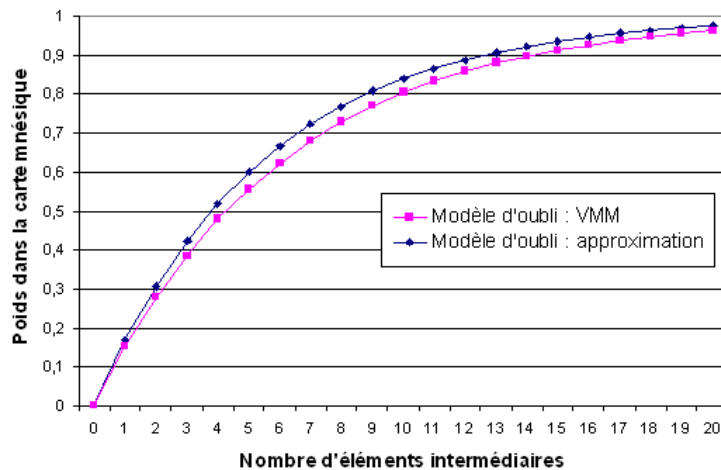


FIGURE 2.20 – Courbes définissant selon le nombre d'éléments intermédiaires le poids en mémoire de l'élément.

est intéressant le modèle va augmenter ses chances d'être refixé, un mot intéressant nécessitant un traitement cognitif plus important. Ainsi au lieu d'avoir la même probabilité d'encodage pour tous les mots, leur poids sémantique est déterminant dans leur encodage. Si le mot est très proche de la question son encodage aura une probabilité légèrement plus faible. La modulation de cette probabilité est telle que la moyenne d'encodage soit la même qu'avant la modification.

Modulation de la mémoire par le sémantique dans le traitement des paragraphes

Pour les documents structurés en paragraphes, nous prenons en compte l'intérêt d'un élément pour son encodage en mémoire. Cette fois-ci nous considérons que tous les paragraphes traités sont encodés, mais que leur seuil initial est dépendant de leur similarité avec le thème. Une séquence de mots jugée intéressante a donc une valeur initiale égale à sa similarité sémantique avec le but de la recherche (voir quelques exemples figure 2.21 page suivante). Par exemple un bloc intéressant est d'ores et déjà un candidat potentiel pour une revisite, au même titre qu'un élément inintéressant mais vu quatre étapes auparavant et donc oublié.

2.6.4 Discussion

Après avoir vu les différentes parties du modèle, intéressons-nous maintenant à sa validation par comparaison avec des données empiriques. Pour cela, plusieurs expériences ont été conçues avec enregistrement des mouvements oculaires des participants. Nous avons fait plusieurs expérimentations afin de tester les différentes parties du modèle, et pour le complexifier au fur et à mesure. Dans toutes ces expérimentations il s'agit pour les participants de rechercher une information non explicite (le but étant de répondre le plus précisément possible à une question).

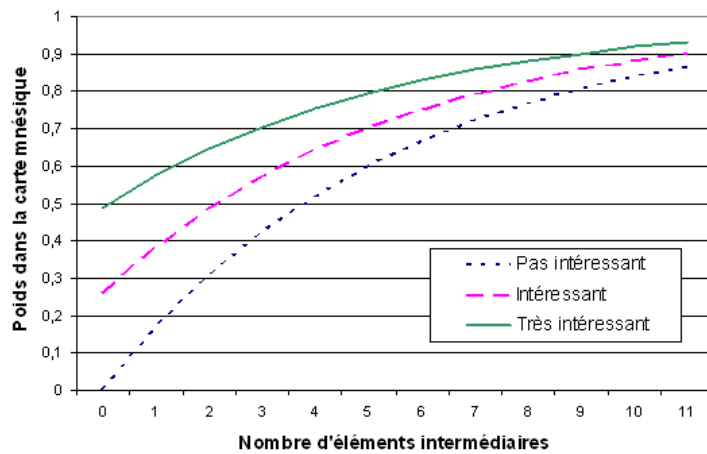


FIGURE 2.21 – Courbes définissant selon le nombre de blocs intermédiaires le poids en mémoire du paragraphe. La valeur initiale égale à la similarité sémantique du texte traité avec le thème de la recherche.

Nous commencerons tout d'abord par une expérimentation assez simple de recherche d'un mot isolé, dont deux versions ont été mises au point. La suite des expérimentations sera plus écologique (plus proche de pages Web) avec des paragraphes organisés de manière structurée (en lignes et en colonnes ou dans une structure plus complexe) dans un document.

3

Application du modèle à un environnement simple

Afin d'étudier le parcours d'éléments ainsi que les contraintes respectives des processus impliqués dans une tâche simple de recherche d'information nous avons mis au point une première expérimentation. Celle-ci est basée sur une page ne contenant que des éléments isolés, des mots ne formant ni phrase ni paragraphe. Le but de la tâche est de trouver parmi quarante mots celui qui répond le mieux à une question. Cette tâche a été choisie de manière à ne pas entraîner de phénomène de reconnaissance simple du mot (*pattern matching*), et pour impliquer un traitement cognitif sémantique des mots traités.

Les mots sont organisés sur la page selon leurs caractéristiques sémantiques et visuelles. En effet, chaque mot a une similarité sémantique avec la question, et une saillance visuelle. Un mot dont la taille de police de caractères est plus élevée ou un mot en couleur se verra affecté d'une « valeur visuelle » plus élevée qu'un mot de taille de police de caractères réduite ou en noir.

Cette première expérimentation a été reproduite selon deux modalités visuelles différentes afin de tester la robustesse du modèle. Dans un premier cas, les mots étaient tous en noir et avaient des tailles variables et dans l'autre cas les tailles étaient fixes, et les mots pouvaient être soit noir soit rouge.

Dans ce chapitre, nous commencerons par la description de la méthodologie expérimentale qui a été mise en place. Suivront les résultats obtenus par les participants, puis une description du modèle appliqué à cette tâche. Enfin nous comparerons les données empiriques aux données simulées par le modèle.

3.1 Expérimentations avec des mots

3.1.1 Méthode

Stimuli

Nos stimuli sont des pages contenant quarante mots isolés (voir figure 3.1 page ci-contre). Ces mots ont deux attributs, un visuel et un sémantique. Ces deux types d'information vont guider la recherche. Pour l'utilisateur, l'objectif est de trouver parmi ces mots celui qui répond le mieux à une question du type : *Quel est le X le plus Y ?*, par exemple : le sport le plus brutal, l'aliment le plus gras, l'animal le plus gros, etc.

Dans la première version de l'expérience, l'attribut visuel des mots est la taille de la police de caractères, variant graduellement de 13pt à 19pt (soit sept tailles). Dans la seconde version, cet attribut visuel est la couleur des mots, avec cette fois-ci deux valeurs : rouge et noir.

Nous avons choisi ces deux types d'attribut visuel afin de comparer les comportements des participants et du modèle en situation multivaluée (avec les tailles) et dichotomique (avec la couleur). L'attribut *Taille* permet de graduellement augmenter la saillance des éléments. Nous avons choisi les tailles de 13pt à 19pt pour des raisons théoriques et pratiques. Ces tailles sont celles que nous trouvons habituellement dans les documents, ou dans les nuages de mots sur le Web (*tag-clouds*). Cet intervalle permet aux quarante mots d'être lisibles à une même distance et de tenir dans une même page sans se chevaucher.

Un effet graduel contrôlé ne peut pas être obtenu avec différentes couleurs car les couleurs ne sont pas ordonnables en terme de saillance. Nous avons donc choisi une seule couleur, saillante, le rouge. L'utilisation de cette couleur n'est pas anodine, car dans notre vie quotidienne le rouge est souvent associé à une alerte pour des informations importantes, interdites, à remarquer. La saillance des éléments rouges est donc plus importante que celle des noirs.

Pour chaque mot, son attribut sémantique correspond à sa similarité sémantique avec la question. Les similarités ont été calculées à l'aide de LSA entraîné sur un corpus français de vingt-quatre millions de mots, composé de tous les articles du quotidien *Le Monde* en 1999. Après avoir imaginé dix-huit questions, sous la forme *Quel est le X le plus Y ?*, nous avons trouvé pour chacune quarante noms communs distribués le plus linéairement possible entre la similarité maximale (1) et la similarité minimale (0) (voir en annexe A page 137 toutes les définitions et tous les mots). Grâce à l'utilisation de LSA, ces listes de mots ont été générées automatiquement, à partir de la question. Ces mots devaient en plus répondre à deux critères : avoir un indice de fréquence compris entre 0,3 et 0,7 afin de n'être ni très fréquents comme les articles, ni très rares risquant ainsi d'entraîner un traitement sémantique différent lors de leur lecture (dans LSA cette valeur n'est pas exactement la fréquence dans le corpus, mais un indice corrélé à la fréquence). Le second critère contrôlé pour ces mots est leur longueur, avec comme minimum cinq lettres et comme maximum neuf lettres,

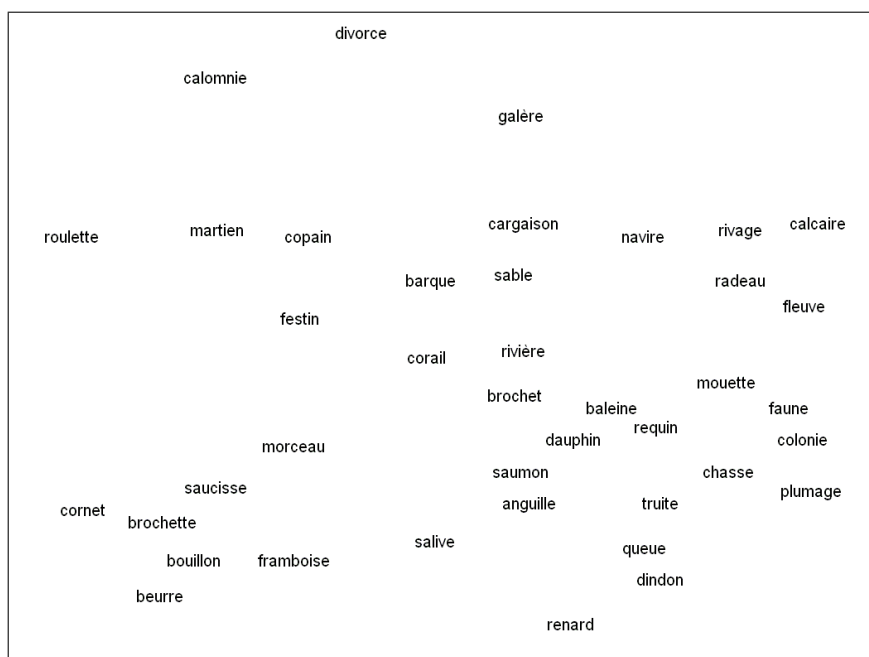


FIGURE 3.1 – Exemple de page vue par les participants. La question était *Quel est le poisson le plus dangereux ?* Pour ce groupe, tous les mots avaient la même taille.

ceci afin de limiter les effets de saillance perceptive non contrôlés. Dans chaque liste, sept mots (dont la cible) appartiennent à la catégorie de la question, les trente-trois autres sont des distracteurs et ont des similarités décroissantes avec la question. Voici par exemple, les mots en lien avec la question *Quelle est la plante la plus piquante ?* (les similarités avec la question sont notées entre parenthèses) :

- | | | |
|--------------------|--------------------|--------------------|
| – cactus (0,71) | – astuce (0,39) | – tisane (0,18) |
| – cerisier (0,63) | – terrasse (0,37) | – cylindre (0,17) |
| – haricot (0,59) | – femelle (0,35) | – promeneur (0,16) |
| – oranger (0,58) | – fumier (0,34) | – brosse (0,14) |
| – menthe (0,57) | – odeur (0,33) | – froid (0,13) |
| – palmier (0,55) | – gousse (0,31) | – fourche (0,11) |
| – pommier (0,53) | – abeille (0,30) | – crapaud (0,10) |
| – graine (0,52) | – feutre (0,28) | – dosage (0,08) |
| – jardinage (0,48) | – finesse (0,25) | – pavillon (0,07) |
| – botanique (0,47) | – rosée (0,24) | – poltron (0,05) |
| – feuille (0,46) | – carnivore (0,23) | – bassin (0,03) |
| – papillon (0,46) | – produit (0,21) | – pollution (0,01) |
| – taille (0,42) | – parchemin (0,20) | |
| – mètre (0,39) | – bocal (0,19) | |

Prenons un autre exemple. Les mots en lien avec la question *Quel est le poisson le plus dangereux ?* sont les suivants (les similarités avec la question sont notées entre parenthèses) :

– baleine (0,73)	– navire (0,44)
– anguille (0,67)	– rivage (0,42)
– requin (0,65)	– barque (0,40)
– truite (0,60)	...
– saumon (0,54)	– copain (.07)
– dauphin (0,53)	– festin (.05)
– brochet (0,45)	– calomnie (.04)
– mouette (0,44)	– roulette (.02)

Un questionnaire auprès de vingt-six personnes (étudiants de l'université de Grenoble pour la plupart) nous a permis de classer plus finement les sept mots de la catégorie selon leur similarité avec la question (voir le questionnaire en annexe B page 142). En effet, LSA n'est pas assez fin pour distinguer toutes les nuances de notre langage, en particulier lorsque les documents sont très courts comme c'est le cas ici (voir section 2.4.1 page 39). Nous ne pouvons pas nous baser sur sa capacité, par exemple, à déterminer quel est le poisson le plus dangereux entre une baleine et un requin. Les participants à ce questionnaire devaient donner à chaque mot un score entre 1 et 5 notant ainsi sa pertinence par rapport à la question.

Les résultats nous donnent une moyenne des mots cible de 4,76 sur 5, et une moyenne des autres mots de la catégorie de 2,31. Ce questionnaire nous a aussi permis de détecter quatre questions pour lesquelles le mot-cible ne faisait pas l'unanimité. Par exemple, à la question *Quel est l'instrument de musique le plus encombrant ?* les mots « piano » et « harpe » ne se départageaient pas (moyennes : 4,85 et 4,81, $T(25)=0,57$, $p=0,57$). Dans ce cas le remplacement de « piano » par « orgue » a résolu ce problème. À partir des scores nous avons reclassé les mots du plus au moins proche de la question. Une fois les mots reclassés de 1 à 7, nous avons redistribué les valeurs de similarités de ces sept mots entre 1 et le cosinus du huitième mot. Le mot-cible ayant la valeur la plus proche de 1. Nous continuerons par la suite à parler de cosinus sachant que pour ces sept mots ce ne sont pas les valeurs de similarités retournées par LSA mais celles recalculées suite aux résultats du questionnaire. Ces cosinus sont représentés avant et après correction sur la figure 3.2 page ci-contre. Une fois les mots choisis, la génération des pages s'est faite en plusieurs étapes.

Les mots ont été placés sur la page de manière à respecter les similarités sémantiques des mots entre eux. Nous avons voulu ici avoir une organisation sémantique de l'espace, les mots proches sémantiquement devant l'être spatialement. Pour la représentation en deux dimensions des mots, nous avons tout d'abord généré la matrice de similarités entre chacun des quarante mots, similarités calculées avec LSA (un exemple de matrice de similarités se trouve en annexe C page 145).

Une fois cette matrice générée, différentes étapes sont nécessaires (illustration figure 3.3 page 70).

1. Un algorithme de *MultiDimensional Scaling* (MDS) a été utilisé afin d'obtenir un placement sur deux dimensions spatiales (en haut à gauche sur la figure).
2. Le résultat obtenu n'était pas utilisable tel quel, il a donc fallu normaliser les

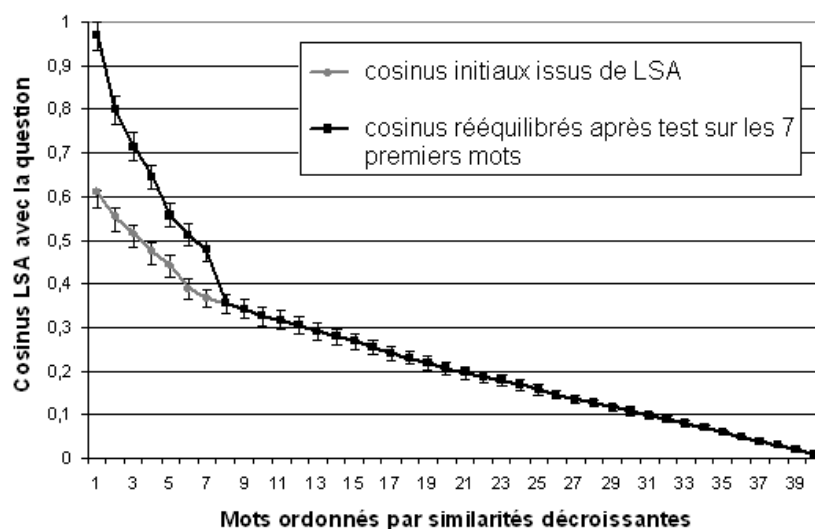


FIGURE 3.2 – Distribution des mots de toutes les pages selon leur similarité à la question à laquelle ils sont associés.

positions des mots afin qu'ils se distribuent dans toute la page.

3. L'étape suivante a permis de les remettre dans un nouvel espace afin de diminuer les effets de groupement et d'éviter que certains mots ne se superposent. À partir d'une grille de 11×13 positions possibles, quatre-vingt positions ont été sélectionnées. La carte uniforme a ensuite été légèrement biaisée avec des variations jusqu'à 3% en vertical et 6% en horizontal afin que la grille ne soit plus clairement visible tout en ayant toujours les points à une certaine distance les uns des autres. Chaque mot a ensuite été rattaché à un point. L'appariement se fait en minimisant la distance totale entre les emplacements de départ et d'arrivée. Sur la figure en bas à gauche, les mots sont les points (en rouge), les cercles sont les quatre-vingt emplacement possibles (dont quarante seront finalement utilisés) et les cercles avec une croix sont les positions finales des mots.
4. Une dernière étape consiste à faire des transformations par symétrie horizontale ou verticale des pages afin d'avoir une répartition uniforme des mots dans toutes les pages et éviter par exemple d'avoir toujours un groupement de mots en bas à gauche.

Dix-huit questions différentes ont été définies (voir les pages en annexe D page 147). Les mêmes questions, et les mêmes mots ont été utilisés pour les deux versions de l'expérience (*Taille et Couleur*)

Variables indépendantes

En vue de tester le modèle, et sa robustesse, nous avons donc réalisé deux versions de cette expérimentation : une version avec des mots de différentes tailles, et une

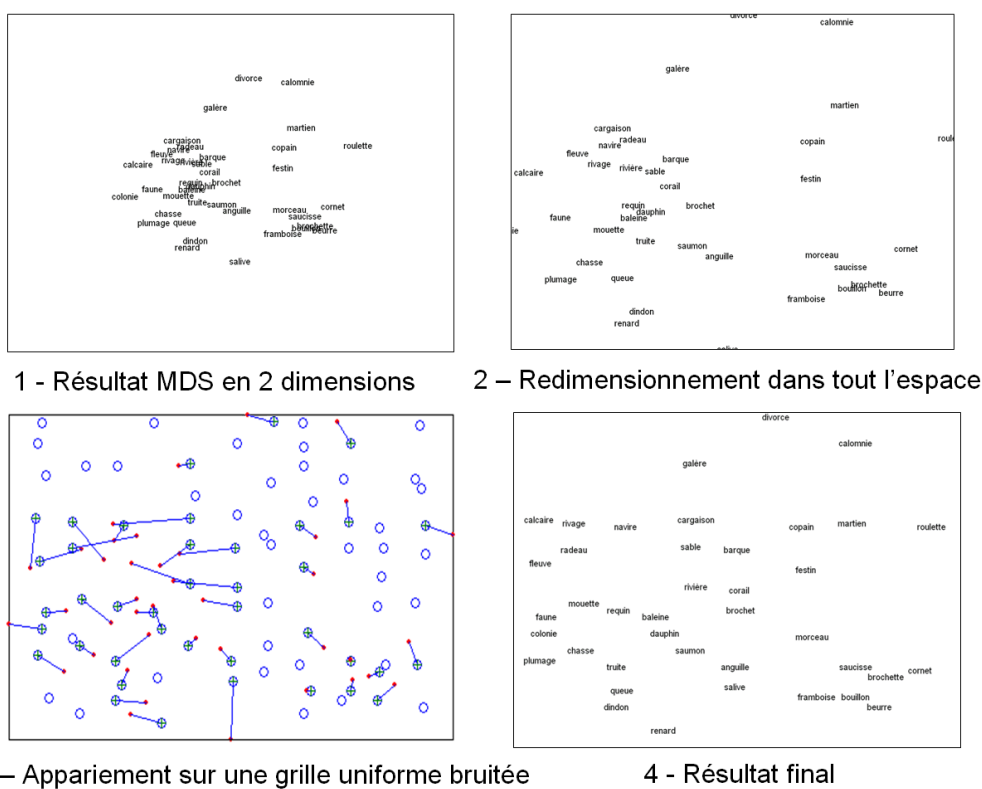


FIGURE 3.3 – Étapes de génération des pages à partir de la matrice de similarités. (1) MDS sur la matrice et sélection des deux premières dimensions ; (2) redimensionnement dans tout l'espace ; (3) distribution sur une grille uniforme ; (4) résultat final.

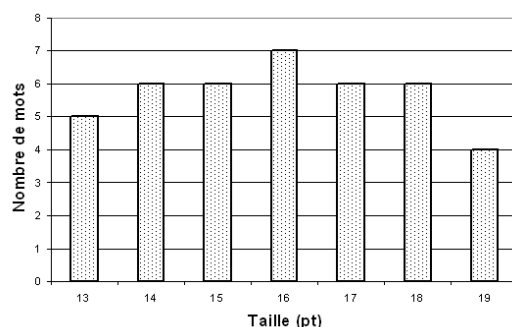


FIGURE 3.4 – Distribution des quarante mots selon leur taille.

version avec des mots de deux couleurs différentes. Dans les deux versions, nous avons choisi deux variables indépendantes, l’organisation sémantique de la page et l’organisation visuelle de la page. La variable visuelle est intra-groupe (un même participant voit les différentes modalités visuelles) et la variable sémantique, inter-groupe (un participant ne voit qu’une seule des modalités).

Dans la première version de l’expérimentation il y a trois modalités visuelles croisées à deux modalités sémantiques. Les modalités visuelles déterminent la taille des mots de manière linéaire et sont les suivantes :

- V+ : congruent avec la localisation des mots, c’est-à-dire que le mot cible est de taille 19 et en s’en éloignant spatialement les mots sont plus petits.
- V= : tous les mots sont de taille constante fixée à 15.
- V- : la taille des mots est aléatoire, mais avec la même répartition du nombre de mots dans les sept tailles différentes qu’en V+.

Dans les modalités V+ et V- la distribution des tailles est répartie de manière quasi-uniforme (voir figure 3.4)¹.

Au niveau sémantique, les deux modalités sont :

- Sémantique : les mots les plus proches sémantiquement le sont aussi spatialement.
- Non sémantique : les mots sont répartis aléatoirement sur les mêmes emplacements pour ne plus avoir l’effet sémantique tout en gardant le même agencement dans l’espace.

Ces différentes modalités nous donnent six pages pour les mêmes mots en croisant les deux variables (un exemple se trouve en annexe E page 151). L’une de ces six pages est très facilitatrice, avec les organisations sémantiques et visuelles congruentes (SemV+), alors que la page NoSemV- est la plus perturbatrice. Dans la page SemV= seul le sémantique aide, dans la page NoSemV+ seul le visuel aide, etc.

Pour la seconde version de l’expérimentation, il n’y a plus que deux valeurs possibles attributs visuels pour les mots au lieu des sept valeurs de tailles de police. Ici chaque mot est rouge ou noir, la taille de caractères est fixe et fixée à 15. La variable couleur se découpe cette fois en deux modalités de présentation (voir figure 3.5), car

1. Il y a moins de mots de taille 19 pour éviter les superpositions de mots

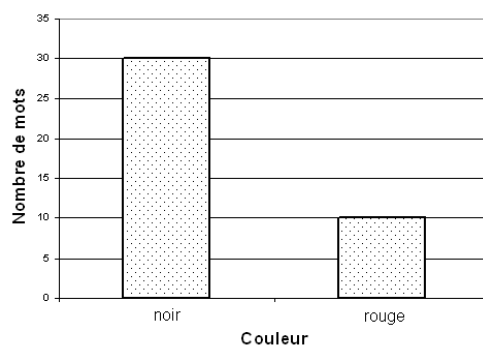


FIGURE 3.5 – Distribution des quarante mots selon leur couleur.

la modalité « V= » aurait été la même que celle de l'expérience précédente :

- V+ : congruent avec la localisation des mots, c'est-à-dire que le mot cible et ses neuf voisins spatiaux sont en rouge et les 30 autres en noir.
- V- : la couleur des mots est aléatoire, mais il y a toujours 10 mots rouges et 30 mots noirs.

Les modalités sémantiques sont identiques à celles de la 1ère version, ce qui nous donne pour une même question quatre pages (voir en annexe E page 151).

3.1.2 Déroulement de l'expérience

Pour les deux versions de l'expérience, un essai commence par la présentation de la définition, puis une croix de fixation, afin de stabiliser le regard, suivie de la page contenant les quarante mots (voir figure 3.6 page suivante). Une fois la réponse trouvée les participants appuient sur une touche et donnent leur réponse oralement à l'expérimentateur. Il n'y a pas de contrainte de temps pour répondre.

Les mouvements oculaires des participants ont été enregistrés avec un oculomètre Eyelink 2, ayant une résolution temporelle de 4 ms, soit une fréquence d'échantillonnage de 250 Hz. Les pages ont été présentées sur un écran 19 pouces, (résolution 1024 × 768), les participants étaient à une distance fixe de 50 cm de l'écran, une mentonnière était utilisée pour restreindre les mouvements de tête des participants. À cette distance, l'écran couvrait environ 40 degrés horizontalement et 30 degrés verticalement. Seul l'œil directeur a été utilisé pour les analyses.

Pour éviter le biais de centralité souvent observé dans la littérature, les positions des croix de fixation ont été réparties suivant les pages à l'un des quatre angles de l'écran, et ce dans une région de la page ne contenant pas ou peu de mots, pour que l'œil ne soit pas directement sur un mot dès le début de la recherche. Ainsi sur les dix-huit pages, quatre ont eu leur croix de fixation en bas à gauche, cinq en bas à droite, cinq en haut à gauche et quatre en haut à droite. La position des croix de fixation était à 100 pixels de chaque bord. Cette croix de fixation était présentée durant 1500 ms puis les participants devaient la fixer pendant 50 ms dans un carré de 100 × 100 pixels pour que le stimulus apparaisse.

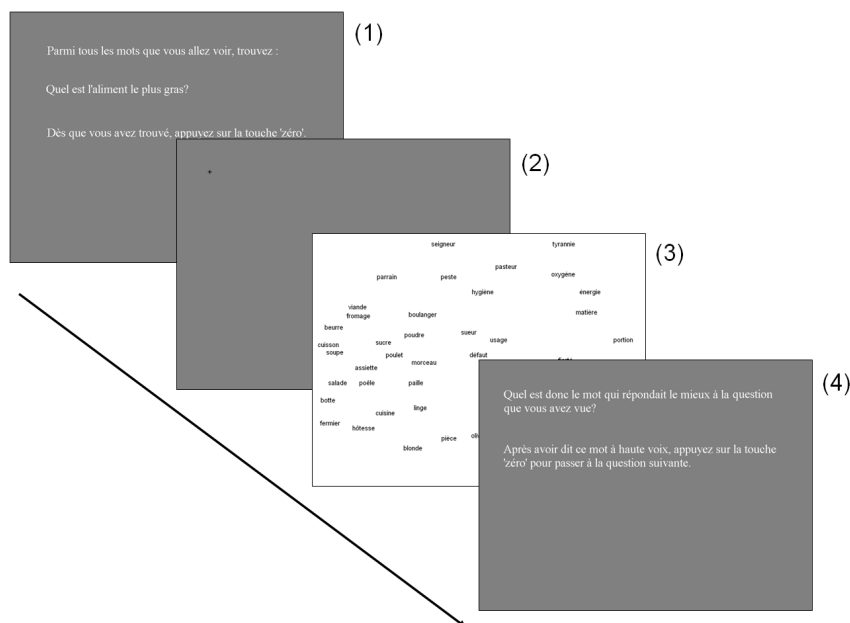


FIGURE 3.6 – Étapes du processus expérimental. Les participants (1) lisent la question ; (2) fixent le point de fixation ; (3) cherchent l'information dans la page avec les quarante mots (4) donnent leur réponse. Exemple pour la question *Quel est l'aliment le plus gras ?*.

Participants

Quarante-trois étudiants (âge moyen : 20,9 ans) ont participé à l'expérience dans sa première version, se répartissant entre les modalités *Sémantique* (vingt et un participants) et *Non Sémantique* (vingt-deux participants). Chaque participant a vu les dix-huit pages, six en modalité visuel facilitateur (V+), six en visuel neutre (V=) et six en visuel perturbateur (V-). Les participants ont été répartis en six groupes afin que toutes les pages soient vues dans toutes les modalités.

Dans la seconde version, vingt-neuf étudiants (âge moyen : 21,2 ans) ont participé à l'expérimentation, quinze en modalité *Sémantique* et quatorze en modalité *Non Sémantique*, chacun ayant vu neuf pages en modalité V+ et neuf en modalité V-.

Chacun des participants avait une vue normale ou corrigée et aucun ne connaissait l'objectif de l'expérimentation. Quelques exemples de trajets oculomoteurs ainsi enregistrés sur différentes pages se trouvent figures 3.7, 3.8, 3.9.

3.1.3 Résultats expérimentaux

Traitement des données

Pour pouvoir comparer ces données entre elles et ensuite avec le modèle, nous avons dû effectuer des pré-traitements sur les données brutes afin de les rendre exploitables. Nous les avons donc épurées pour garder non plus une suite de coordonnées (X, Y) , mais une suite de mots, le modèle effectuant un trajet de mot en mot.

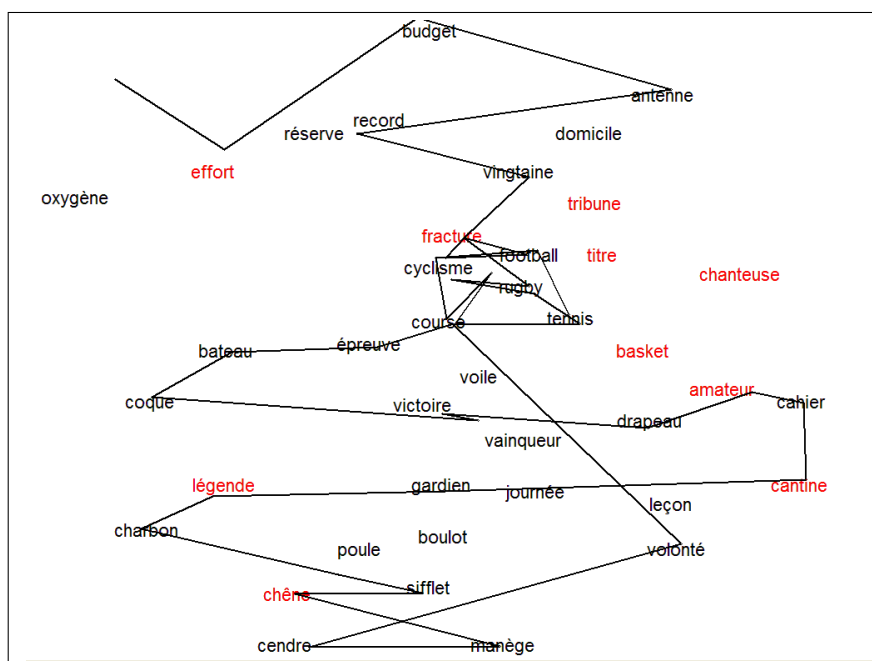


FIGURE 3.7 – Exemple de trajet expérimental pour la question *Quel est le sport le plus brutal ?* dans la condition *Sémantique V-* de l'expérience *Couleur*.

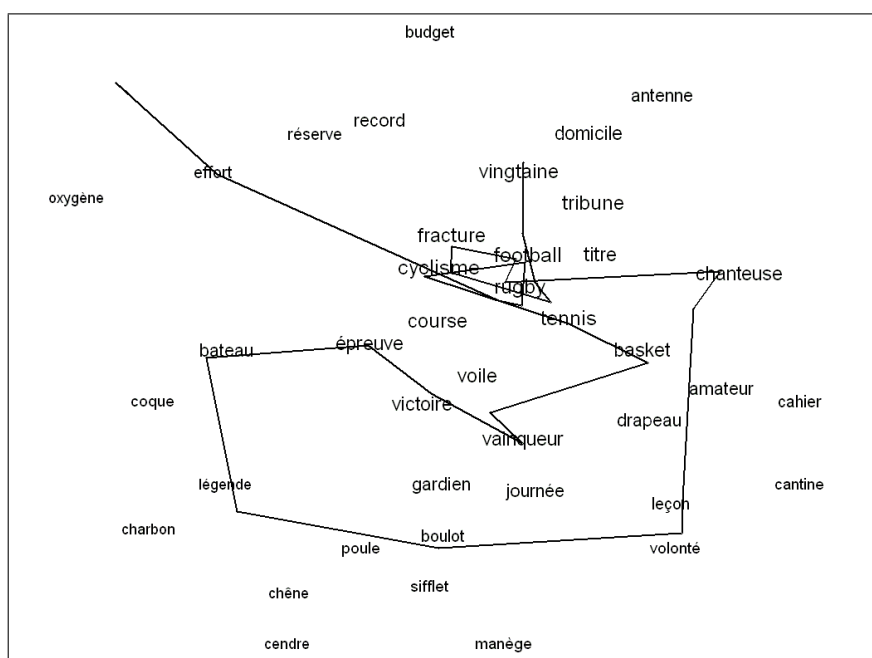


FIGURE 3.8 – Exemple de trajet expérimental pour la question *Quel est le sport le plus brutal ?* dans la condition *Sémantique V+* de l'expérience *Taille*.

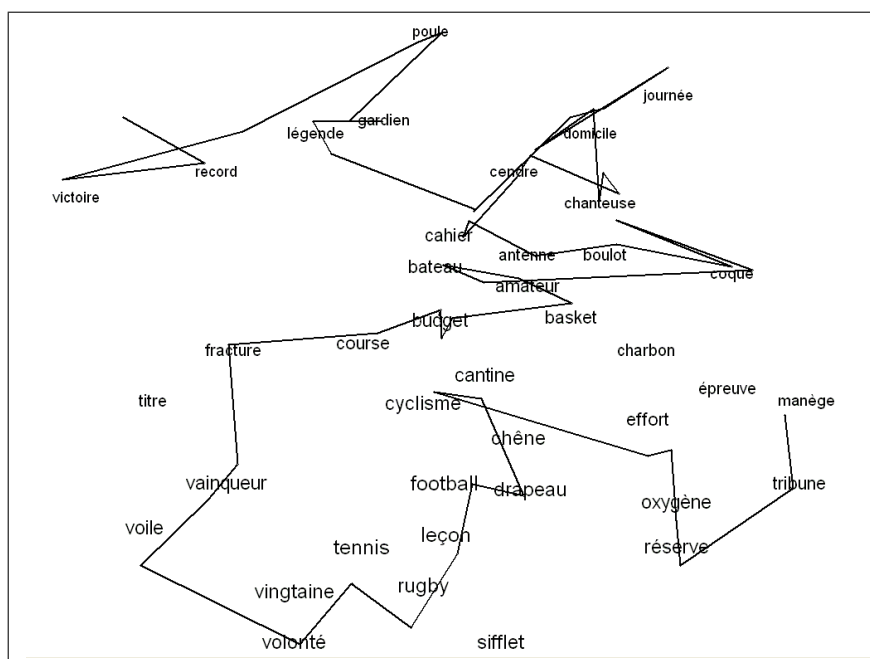


FIGURE 3.9 – Exemple de trajet expérimental pour la question *Quel est le sport le plus brutal ?* dans la condition *Non Sémantique V+* de l'expérience *Taille*.

Nous avons associé à chaque fixation le mot le plus proche, puis supprimé les doublons ce qui nous donne des trajets de mot en mot. À partir de ces données enrichies et épurées, nous avons pu observer des différences de comportement grâce à plusieurs indices.

Dans l'intention de comparer les trajets empiriques avec ceux simulés, nous avons dû trouver des variables globales de haut niveau permettant de décrire ces trajets. Les trajets ne sont en effet pas comparables directement, car il y a une trop grande variabilité inter-individuelle. Par contre, cette comparaison pourra se faire indirectement à travers des variables descriptives de ces trajets. À l'aide de ces variables, nous pouvons décrire des tendances et ainsi comparer les données expérimentales et simulées.

Dans l'étude des mouvements oculaires plusieurs variables sont classiquement utilisées, les cartes de densité, les durées de fixations dans des zones d'intérêt, le nombre de fixations, la durée de première fixation, l'amplitude des saccades, les transitions entre zones d'intérêt. En lecture, la position des fixations dans le mot est utilisée ainsi que la durée de regard (fixation + refixation), etc. Dans notre étude, nous avons analysé les données de manière adaptée à nos stimuli et à notre tâche, et nous avons mis en place une méthodologie adaptée à notre contexte.

Nous avons mis en place deux types de variable descriptive des trajets à des fins différentes de comparaison : les *indices* et les *distributions*. Dans la terminologie de l'apprentissage automatique, les *indices* servent à l'apprentissage du modèle et les *distributions* servent de test. D'une part, pour comparer les modalités entre elles, pour mettre en évidence un effet possible d'apprentissage au cours de la passation et aussi

pour paramétrer le modèle nous avons choisis deux *indices* caractéristiques devant être modulés par les modalités de passation. Ces deux *indices* sont le *nombre de mots vus avant la cible* et le *taux de saccades progressives*. D'autre part les comparaisons entre les données expérimentales et simulées se feront au travers des *distributions* de variables liées au déploiement d'un trajet oculomoteur, c'est-à-dire les *distributions* des angles dans le trajet (absolus et relatifs) et la distribution des amplitudes de saccades.

Apprentissage

Les analyses présentées dans ce paragraphe ont pour but de caractériser les trajets oculomoteurs empiriques pour les comparer ensuite avec ceux du modèle. Nous nous sommes donc particulièrement intéressés à la modalité *Sémantique*. C'est effectivement dans cette modalité que tous les processus entrent en interaction. C'est donc sur ces données dans la modalité *Sémantique* que le modèle sera construit. Le modèle ne possède pas de conditions d'arrêt mais est programmé pour effectuer autant de fixations que les participants sur une carte. Nous avons mesuré le nombre de mots vus avant d'atteindre la cible comme indice de la performance des participants.

Cet indice nous a permis de mettre en évidence un phénomène d'apprentissage de l'organisation sémantique des mots. En effet, entre les premières et les dernières pages vues pendant la passation, les performances des participants dans la modalité *Sémantique* s'améliorent significativement, ($T(65) = 2,44, p < 0,02$), voir figure 3.10 page ci-contre. La séparation de l'expérience en trois périodes nous montre un apprentissage de la répartition sémantique des mots, entraînant une réduction du nombre de mots vus pour atteindre la cible. Afin de confirmer cet effet d'apprentissage, nous avons calculé le taux de saccades progressives, c'est-à-dire le pourcentage de saccades se rapprochant de la cible avant d'atteindre la cible pour la première fois. Un pourcentage élevé dénote un trajet direct vers la cible alors qu'un pourcentage faible correspond à un trajet « hésitant ». L'apprentissage est significatif ($T(115) = 4,42, p < 0,01$) comme nous pouvons le voir sur la figure 3.11 page suivante, cette séparation de l'expérience en trois périodes nous montre un apprentissage de la répartition sémantique des mots, entraînant un rapprochement plus franc vers la cible.

Pour comparer les données expérimentales à celles du modèle nous utiliserons donc les trajets effectués dans la dernière phase expérimentale (pages 13 à 18) dans la modalité *Sémantique*, une fois que les participants ont bien compris la tâche et l'organisation sémantique des mots sur l'écran.

Distributions de caractéristiques

Un bon modèle doit pouvoir produire un trajet ayant une géométrie proche de celle des humains avec des amplitudes de saccades également similaires.

Pour caractériser la forme générale des trajets, nous avons utilisé deux distribu-

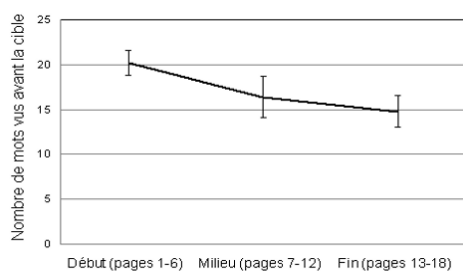


FIGURE 3.10 – Nombre de mots vus pour les participants de l’expérience (version *Taille*) dans la modalité *Sémantique*.

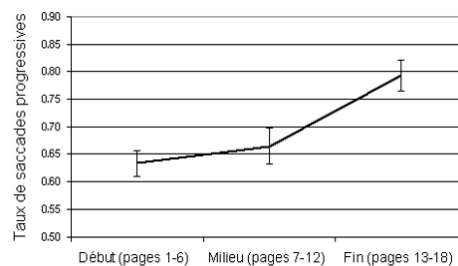


FIGURE 3.11 – Pourcentage de saccades se rapprochant de la cible dans les trajets des participants de l’expérience (version *Taille*) dans la modalité *Sémantique*.

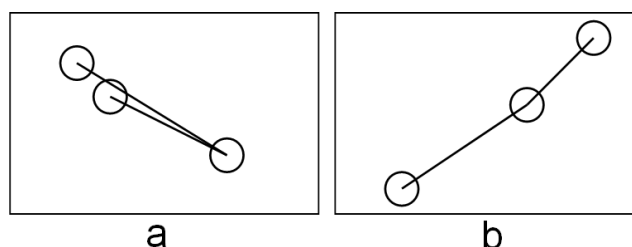


FIGURE 3.12 – (a) Angle relatif proche de 0° (demi-tour). (b) Angle relatif proche de 180° (tout droit).

tions des angles : les angles absolus des saccades par rapport à l’horizontal, c’est-à-dire leur direction dans l’espace (N, E, S, O), et les angles relatifs, c’est-à-dire l’angle formé par la saccade courante et la saccade précédente, entre 0° (demi-tour) et 180° (tout droit) (voir schéma figure 3.12).

Nous avons aussi pris en compte la distribution des amplitudes des saccades (en degré d’angle visuel). Ces trois distributions sont très proches pour les deux versions de l’expérience, comme nous pouvons le voir dans les figures 3.13, 3.14 et 3.15. Sur la première nous pouvons voir que les directions horizontales sont privilégiées par rapport aux directions verticales et ces dernières par rapport aux directions obliques. Dans la seconde, nous observons que la tendance est à continuer tout droit (180°) avec néanmoins un pic sur les demi-tour complet (0°). Enfin la dernière courbe nous montre que les saccades sont assez courtes, pic à 4° et moyenne de $6,13$, sachant que l’espace moyen entre deux mots est de $4,93^\circ$ (de centre à centre). Elles représentent donc les distributions que le modèle devra retrouver.

Comparaison entre les deux versions de l’expérience

Nous avons ici deux versions de l’expérience dont nous voulons comparer les résultats afin de vérifier qu’il y a suffisamment de variabilité entre les deux.

En effet, s’il y a une variabilité entre les comportements le modèle sera bien testé

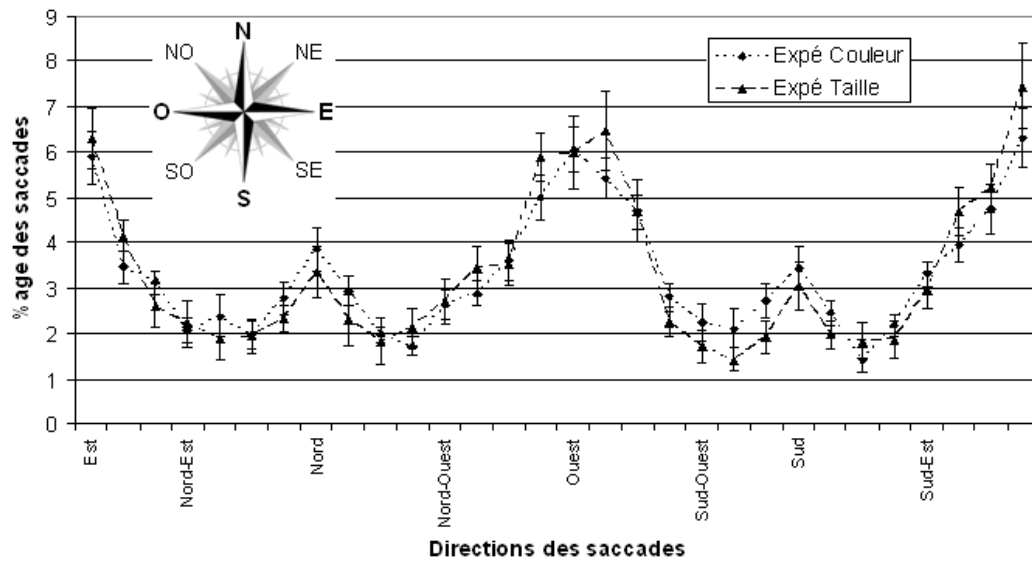


FIGURE 3.13 – Distributions des saccades selon leur direction.

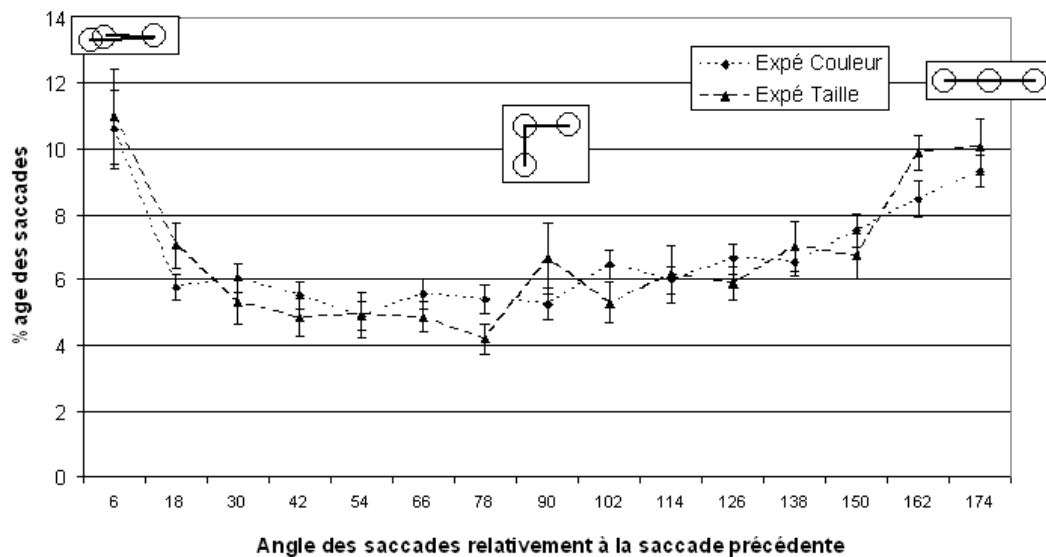


FIGURE 3.14 – Distributions des saccades selon leur cap.

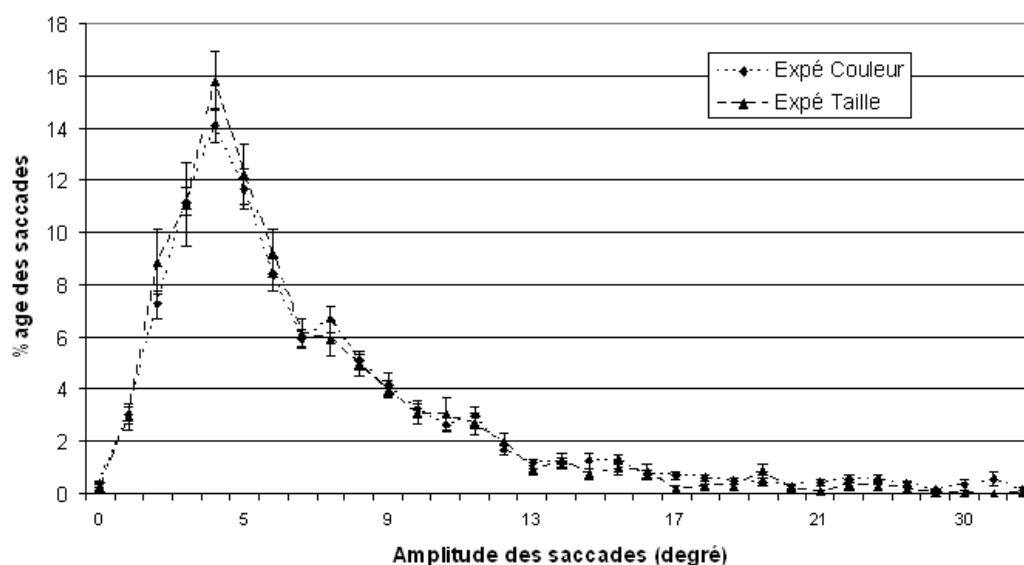


FIGURE 3.15 – Distributions des amplitudes des saccades.

sur les situations différentes, et nous pourrions voir s’il est assez robuste pour simuler correctement des trajets sur ces deux types de stimuli. Dans ces deux expériences, nous avons identifié des indices qui nous permettent de caractériser les performances des participants.

La différence entre les deux expériences est au niveau du facteur visuel, dans un cas les mots sont de tailles différentes avec une évolution graduelle, dans l’autre chaque mot est soit rouge soit noir. Les résultats expérimentaux montrent (voir figures 3.16 et 3.17 page suivante) des gains de performance pour la version *Couleur* par rapport à la version *Taille* entre les modalités V+ et V-.

Les valeurs observées pour *le nombre de mots avant la cible* sont significativement différentes entre les deux modalités visuelles (V+ et V-) pour la version *Couleur* ($T(17) = 2,06, p = 0,055$) mais pas pour la version *Taille* ($T(17) = 0,96, p = 0,349$). La variation dans la taille des éléments n’a donc pas d’influence significative sur les performances, contrairement aux couleurs. Pour la variable *taux de saccades progressives*, nous obtenons un résultat similaire, à savoir une différence significative entre les modalités visuelles lorsqu’il s’agit de couleur ($T(17) = 4,07, p \leq 0,001$) et pas de différence significative quand la variation porte sur les tailles ($T(17) = 0,73, p = 0,474$).

Pour essayer de mieux comprendre d’où viennent ces différences de performance entre les deux expériences, et surtout, pourquoi le facteur *Taille* n’a pas autant d’incidence sur la performance que le facteur *Couleur*, nous avons étudié les fréquences de fixation des mots. Ceci afin de voir si les mots en rouge sont plus fixés que les mots en noir, et de la même façon, si les mots de grande taille (18 et 19) sont plus fixés que les mots plus petits. Pour cette analyse, nous avons pris en compte les pages dans lesquelles les mots n’étaient pas organisés visuellement, c’est-à-dire la modalité V-

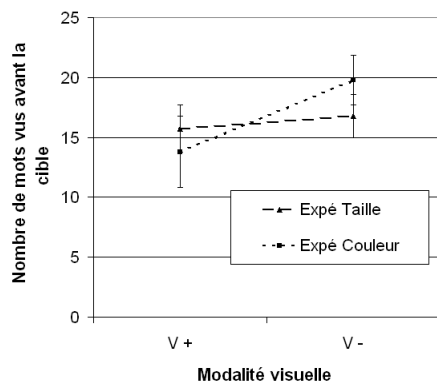


FIGURE 3.16 – Nombre de mots vus avant d’atteindre la cible pour les modalités V+ et V- pour les deux versions de l’expérience. La condition V+ n’est facilitatrice que dans la version *Couleur*.

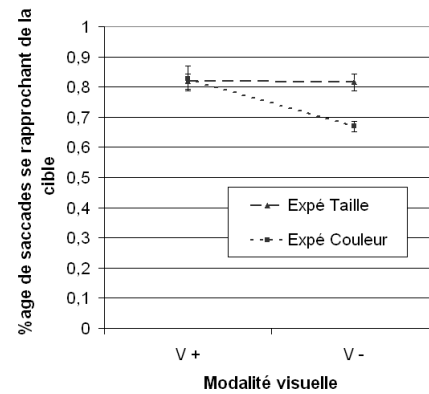


FIGURE 3.17 – Taux de rapprochement vers la cible pour les modalités V+ et V- pour les deux versions de l’expérience. La condition V+ n’est facilitatrice que dans la version *Couleur*.

(en V+ la congruence de la sémantique et du visuel ne permet pas d’isoler le facteur visuel). Les résultats (figure 3.18 page ci-contre) montrent que les mots visuellement saillants ne sont pas plus fixés que les autres. Il n’y a pas de différence notable entre les valeurs observées et théoriques (facteur *Taille* : $\chi^2 = 1,26$, dl = 1, p = 0,26 et facteur *Couleur* : $\chi^2 = 2,36$, dl = 1, p = 0,12). L’accroissement régulier des tailles est donc certainement la raison de la différence observée dans les comportements.

Partant de ces constats nous allons maintenant voir comment le modèle arrive à prendre en compte ces différences de comportements, ainsi que le comportement global des participants. Nous commencerons par un bref rappel des choix de modélisation associés à cette expérience, puis nous verrons la comparaison à proprement parlé et les conclusions que nous pouvons en tirer sur les interactions des processus cognitifs mis en jeu.

3.2 Modélisation et comparaisons avec les données empiriques

3.2.1 Les trois processus et leur interaction

Comme nous l’avons décrit dans le chapitre 2 page 31, des cartes topographiques vont être générées dynamiquement pour chaque processus et combinées afin de prédire le parcours du modèle de mot en mot. Trois processus sont en interaction, visuel, sémantique et mnésique et nous allons grâce à cette première série d’expériences déterminer le poids relatif de ces processus dans cette tâche de recherche d’information. Les documents dans lesquels s’effectue la recherche d’information se présentent sous forme de constellation de mots, répartis selon leur proximité sémantique, ce qui in-

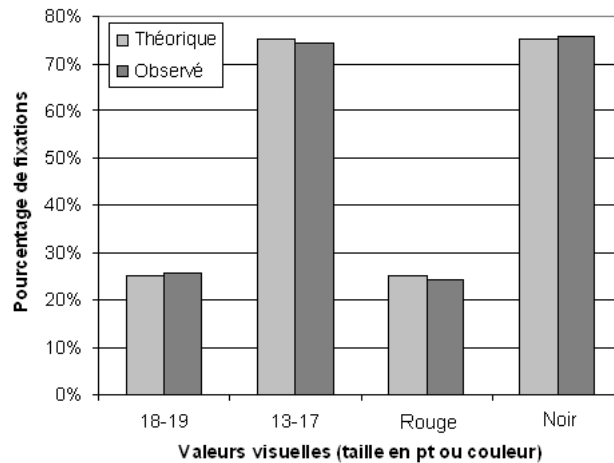


FIGURE 3.18 – Répartition des fixations entre les mots selon leurs caractéristiques visuelles (taille ou couleur), comparée à la répartition théorique.

fluence les choix de modélisation des processus.

Processus Visuel

Au niveau visuel, étant donnée l'organisation de la page, nous avons implémenté un processus visuel local. En effet, il n'y a pas de structure visuelle évidente (ligne, colonne, etc.) la seule organisation étant d'ordre sémantique. Le processus visuel se base sur la saillance visuelle des éléments, modulée par l'acuité visuelle.

$$poidsVis(mot) = saillanceVis(mot) \times acuitéVis(mot, fixationCourante) \quad (3.1)$$

Dans la version *Taille*, la saillance visuelle tient compte de la surface occupée par le mot, c'est-à-dire le carré de la taille. Une normalisation (division par la taille maximale, 19) permet d'avoir des valeurs comprises entre 0 et 1. Pour la version *Couleur* les éléments en rouge ont un poids Ω et les autres un poids de $1 - \Omega$. La valeur de Ω a été fixée arbitrairement à 0,75. Cette valeur simule le contraste chromatique entre le rouge et le noir, et intègre aussi le fait que le rouge a une valeur informative forte.

$$saillanceVis_{Taille}(mot) = \left(\frac{tailleFont(mot)}{19} \right)^2 \quad (3.2)$$

$$saillanceVis_{Coul}(mot) = \begin{cases} \Omega & \text{si coul}(mot)=\text{rouge} \\ 1 - \Omega & \text{sinon} \end{cases} \quad (3.3)$$

Chaque mot a une valeur selon la position de la fixation courante. Une carte est donc générée à chaque fixation, entraînant en partie le choix de la fixation suivante.

Processus Sémantique

Les mots sont organisés sémantiquement sur l'écran. Cette organisation va donc servir de base à la modélisation du processus sémantique dans cette tâche. Nous avons fait le choix d'associer à chaque mot une valeur dépendant de la fixation courante et de la similarité sémantique du mot courant avec la question.

Si ce mot est pertinent par rapport à la recherche, la zone l'entourant sera rehaussée tandis que si ce mot est éloigné de la recherche, la même zone sera rabaisée. Plus le mot est proche de la question plus cette augmentation sera forte, et au contraire plus le mot est éloigné, plus la diminution sera franche. La limite entre les deux se situe pour une similarité entre le mot et la question à 0,2. Ce seuil représente la neutralité, la zone environnante n'augmentant ni ne diminuant (se référer à la section 2.6.2 page 57).

Comme présentée dans le chapitre 2 l'équation de la fonction gaussienne, une fois la hauteur déterminée est la suivante :

$$poidsSem = 0,5 + (hauteur \times e^{-(2 \times distAFixCourante)^2}) \quad (3.4)$$

Processus mnésique

Au niveau de la mémoire, nous avons implémenté le *Variable Memory Model* en attribuant à chaque instant un poids à tous les mots. Au départ tous les mots ont un poids à 1 (attractivité maximale), indiquant qu'ils n'ont jamais été vu. Si un mot a déjà été vu et encodé son poids sera mis à zéro (aucune attractivité), si au contraire il n'a pas été encodé, il aura de nouveau un poids à 1.

La seconde partie de ce processus entraîne un oubli des mots, leurs poids augmentant progressivement au fur et à mesure du trajet (récupération de l'attractivité). Les courbes de mémoire figurent à la section 2.6.3 page 60. Ce mécanisme est formalisé par l'équation suivante :

$$poidsMem_{i,n} = 1 - \Phi^n \quad (3.5)$$

avec i représentant l'élément, n son rang dans le trajet et Φ la probabilité de récupération du modèle de mémoire.

Interaction des processus

Chaque mot a , à un instant t , un poids pour chaque composante, la combinaison des processus correspond à une somme pondérée de ces poids.

$$C_{générale} = \alpha_V \cdot C_{visuelle} + \alpha_S \cdot C_{sémantique} + \alpha_M \cdot C_{mnésique} \quad (3.6)$$

avec $\alpha_V + \alpha_S + \alpha_M = 1$

La suite du trajet, c'est-à-dire la fixation suivante, est située sur le mot ayant le plus fort poids total. Les pondérations des cartes seront fixées grâce à ces deux premières expérimentations.

3.2.2 Simulations

Maintenant que nous avons revu la composition du modèle, nous allons voir plus pratiquement la réalisation des simulations. Notre modèle n'a pas de condition d'arrêt et une fois le trajet entamé, il ne s'arrête qu'une fois un nombre donné de fixations réalisées. En observant les réalisations expérimentales, nous nous sommes aperçus que cette tâche implique une très grande variabilité de trajets avec plusieurs passages sur le mot cible et les mots de la catégorie. De plus, nous n'avons pas mis en évidence de stratégie d'arrêt commune (certains trajets se finissent sur le mot cible, d'autres sur un mot de la catégorie, et d'autres sur un distracteur). Nous n'avons donc pas pu modéliser avec précision le processus de décision qui permet de stopper la recherche, cette fonctionnalité va au-delà de ce que le modèle peut exprimer. Nous avons fait le choix de déterminer le nombre de fixations du modèle a priori à partir des données empiriques. Pour cela nous avons considéré le nombre moyen de fixations des participants sur chaque page.

L'implémentation du modèle comprend une partie visualisation nous permettant d'avoir un aperçu des trajets du modèle selon les paramètres utilisés. Par exemple pour la question *Quel est le poisson le plus dangereux ?* nous obtenons les cartes figure 3.19 page suivante.

3.2.3 Comparaison des données

Notre modèle comprend trois processus, chacun ayant un rôle particulier. Nous n'avions pas d'a priori sur l'importance de chacun de ces processus par rapport aux autres. Un paramètre est associé à chaque processus modélisant son importance dans la simulation de trajets oculomoteurs. La somme des trois paramètres α_V , α_S et α_M est fixe et égale à 1. Nous avons cherché à estimer quelle était la meilleure combinaison possible de paramètres, c'est-à-dire celle qui permettait d'être le plus proche des données expérimentales.

À ces trois paramètres associés aux processus, s'ajoutent deux autres paramètres, qui sont les probabilités liées au mécanisme de mémoire. Là encore nous n'avions pas d'a priori sur la probabilité d'encodage ou de récupération des éléments, nous avons donc estimé de la même manière ces deux paramètres.

Estimation des paramètres

Pour chaque combinaison de paramètres α_V , α_S et α_M , telle que $\alpha_V + \alpha_S + \alpha_M = 1$ avec un pas de 0,01, et pour toutes les combinaisons de ϕ et θ (paramètres du processus mnésique) dans l'intervalle $[0, 5 - 1]$ avec un pas de 0,01, nous avons exécuté le modèle vingt fois sur chaque page (en modalité *Sémantique*). À partir des trajets ainsi simulés nous avons regardé quels étaient les écarts entre les valeurs des participants et celles du modèles pour sélectionner les meilleurs paramètres du modèle. Nous avons donc comparé les erreurs entre les données empiriques et simulées

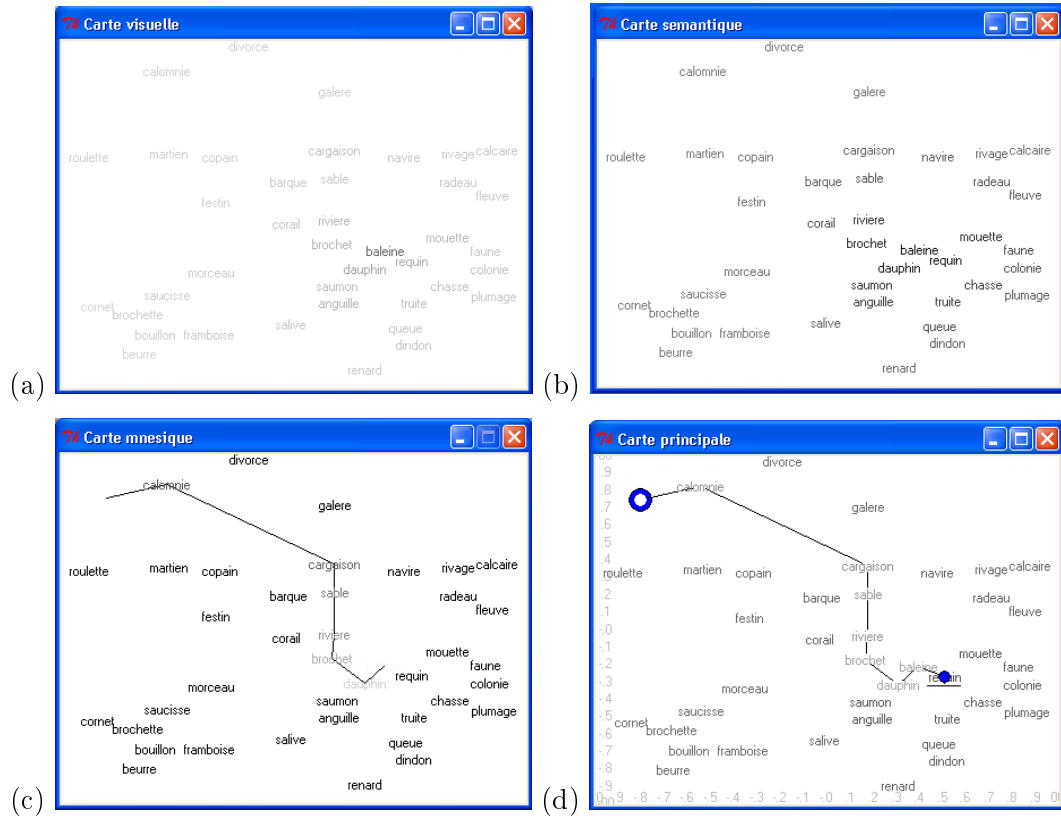






FIGURE 3.19 – Visualisation à un instant t d'un trajet simulé par le modèle, avec comme paramètres $\alpha_V = 0,35$, $\alpha_S = 0,05$ et $\alpha_M = 0,6$. La fixation courante est sur le mot *baleine*. Les mots foncés ont un poids fort et les mots clairs ont un poids faible, n'attirant pas l'attention. (a) Carte visuelle : le mot *baleine* est en zone fovéale avec une très bonne acuité visuelle. Cette acuité diminue autour de ce point, le processus visuel a donc tendance à privilégier la zone proche du point de fixation courant. (b) Carte sémantique : le mot *baleine* est proche de la question, le processus sémantique va donc privilégier la zone proche de ce mot car les mots sont organisés sémantiquement. (c) Carte mnésique : le mot *baleine* étant fixé, le processus mnésique cherche à aller sur un nouveau mot, les mots vu précédemment ont donc une valeur plus faible (en gris) que les nouveaux mots (en noir). (d) Carte intégrative pour la question *Quel est le poisson le plus dangereux ?* à cet instant t . La prochaine fixation a été calculée à partir des trois cartes et se trouve sur le mot *requin*.

TABLE 3.1 – Meilleures combinaisons de paramètres

	Expérience <i>Taille</i>		Expérience <i>Couleur</i>	
	Tx sacc prog.	Nbre mots	Tx sacc prog.	Nbre mots
α_V	0,4	0,34	0,38	0,31
α_M	0,50	0,62	0,45	0,58
α_S	0,11	0,05	0,18	0,12
				
θ	0,83	0,84	0,83	0,85
ϕ	0,84	0,86	0,84	0,86

pour les deux indices décrits auparavant : le nombre de mots vus avant la cible et le taux de saccades progressives.

Nous avons pris en compte pour chaque indice la moyenne des dix meilleures combinaisons de paramètres. Les différences entre les erreurs sont en effet trop faibles pour ne prendre que la meilleure combinaison. Ces résultats sont présentés dans le tableau 3.1.

Nous pouvons tout d'abord remarquer que les meilleurs paramètres sont très proches pour les deux expériences, démontrant que le modèle est adapté pour deux types d'attribut visuel, bimodal dans le cas de la version *Couleur* de l'expérience et progressif avec la version *Taille*. Si maintenant nous interprétons ces valeurs, la plus importante est le poids de la composante mnésique, qui permet au modèle de se rappeler des localisations déjà visitées. La composante visuelle joue également un rôle important, car elle tient compte de l'acuité visuelle et guide donc l'attention sur les mots les plus proches, comme nous l'observons chez l'homme. Enfin, la sémantique est moins importante, en particulier dans la première expérience. Les paramètres θ et ϕ sont également proches de ceux trouvés dans la littérature, qui sont : $\theta = 0,82$ et $\phi = 0,86$ (Horowitz, 2006). Rappelons que l'expérience utilisée dans cette modélisation est nettement différente, ce qui donne une certaine généralité à ces valeurs de paramètres. Il s'agissait d'une recherche visuelle, le but étant de trouver un 2 parmi des 5. Les variables utilisées étaient les temps de réaction.

Validation des paramètres

Les paramètres du modèle ont été estimés à partir de deux indices, le nombre de mots vus avant la cible et le taux de saccades progressives. Afin de valider ces paramètres nous avons utilisé d'autres variables : les distributions des angles (relatifs et absolus) et de l'amplitude des saccades qui caractérisent les trajets empiriques, et que nous avons présentées précédemment.

La figure 3.21 page 88 nous montre cette comparaison pour la première version de

l'expérience (*Taille*), et la figure 3.22 page 89 pour la seconde version de l'expérience (*Couleur*). Nous pouvons noter un bon ajustement des données pour la courbe de distribution des amplitudes des saccades (panneau inférieur des deux figures) avec un pic à environ 4 degrés dans tous les cas. Dans le cas des angles relatifs, les résultats montrent une courbe pour les participants avec beaucoup de saccades régressives (demi-tour, 0°) et de saccades en prolongation avec la précédente (180°), ce qui correspond aux résultats de la littérature (Tatler & Vincent, 2008). Pour le modèle, la courbe est plus plate (voir panneau central dans les figures). Les valeurs que nous obtenons dans le cas des angles absolus sont intéressantes. Elles montrent très clairement une tendance à faire des saccades horizontales (pics à 0° et 180°), plutôt que verticales (les petits pics à 90° et 270°) et ce, pour les participants et le modèle.

Des tests du χ^2 ne nous donnent pas de différences significatives pour chaque comparaison. Les valeurs de p (tous les $p \geq 0,9$) indiquent que les distributions issues du modèle sont proches de celles observées pour ces deux versions de l'expérience.

Différences entre les modalités expérimentales

Nous nous sommes finalement intéressés à la distinction entre les pages avec des mots en rouge autour de la cible (V+) et les pages avec des mots en rouge répartis de façon aléatoire (V-) pour les participants et le modèle. En effet, nous avons vu dans la section sur les résultats des participants une amélioration des performances sur la version *Couleur* lorsque les mots en rouge étaient autour de la cible (voir figures 3.16 et 3.17 page 80). Cette amélioration vient du fait que dans la modalité V+ de la version *Couleur* la localisation des mots proches spatialement de la cible est mieux perçue que dans la version *Taille*. Dans cette version en effet les tailles sont graduelles et le pas d'accroissement des tailles est peut-être trop faible. Les paramètres du modèle sont les mêmes dans ces deux conditions, mais nous observons des différences de performances suivant une même tendance, comme le montre la figure 3.20 page suivante pour le nombre de mots vus avant d'atteindre la cible (différence significative entre les deux modalités pour le modèle, $T(17) = 12,19, p < 0,01$). La modélisation de la prise en compte de la couleur avec le paramètre Ω (fixé à 0,75) permet donc de simuler les différences entre les modalités expérimentales (V+ et V-).

3.3 Discussion

Cette première expérience nous a permis d'établir les bases du modèle. Ce modèle prend en compte à la fois des informations sémantiques et visuelles, et associe à cette double prise d'information un modèle de mémoire visuelle.

Nous avons à déterminer plusieurs paramètres, et pour ce faire, nous avons essayé de faire des choix cognitifs plausibles. La partie visuelle tient compte de l'acuité visuelle humaine et de la saillance des stimuli. Cette saillance est ici très simple en raison de la simplicité des stimuli, mais pourrait être plus complexe si nécessaire. Pour

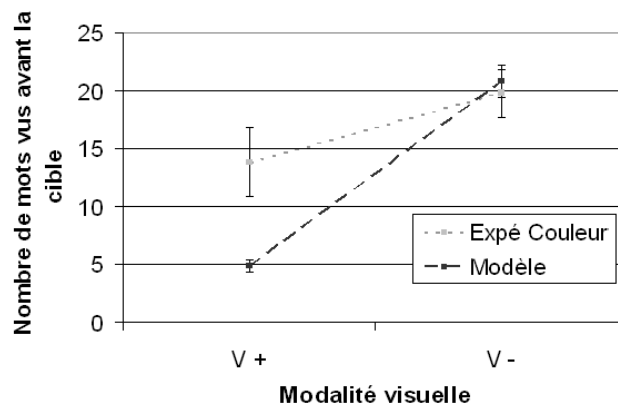


FIGURE 3.20 – Nombre de mots vus avant la cible pour les participants à l’expérience *Couleur* et le modèle.

la partie sémantique, nous avons utilisé une méthode cognitive (LSA) connue pour donner une bonne mesure de similarité entre l’objectif de la recherche d’information et les éléments. Enfin, la partie mémoire, qui est la plus importante selon les résultats du modèle, permet de garder à l’esprit les éléments déjà vu puis de les oublier sans être trop strict, contrairement au mécanisme classique d’inhibition de retour, qui est généralement utilisé dans un tel modèle. Celui-ci est plus réaliste, parce qu’il permet des retours fréquents sur l’item précédemment fixé, comme nous l’avons constaté dans la distribution des angles relatifs.

Nous avons choisi deux échelles différentes pour la saillance des stimuli visuels : une échelle bimodale pour la couleur et une échelle graduelle (sept niveaux) pour la taille. La couleur n’étant pas une caractéristique ordonnable, l’inverse aurait été difficile à maîtriser sur les effets de saillance. Ce choix nous a permis de voir les différences dans les performances, et même si les deux expériences ne sont pas similaires, les meilleurs paramètres pour le modèle sont à peu près les mêmes, ce qui signifie que ce modèle est robuste, au moins face à un tel changement. Cette première étape de validation du modèle a fait l’objet d’une publication (Chanceaux, Guérin-Dugué, Lemaire, & Baccino, 2009a).

D’autres expériences avec des stimuli plus réalistes et se rapprochant de la construction de pages web (journaux en ligne) avec des paragraphes de texte au lieu de simples mots, nous permettront de tester de nouveau ce modèle, comme nous allons le voir dans les chapitres suivants.

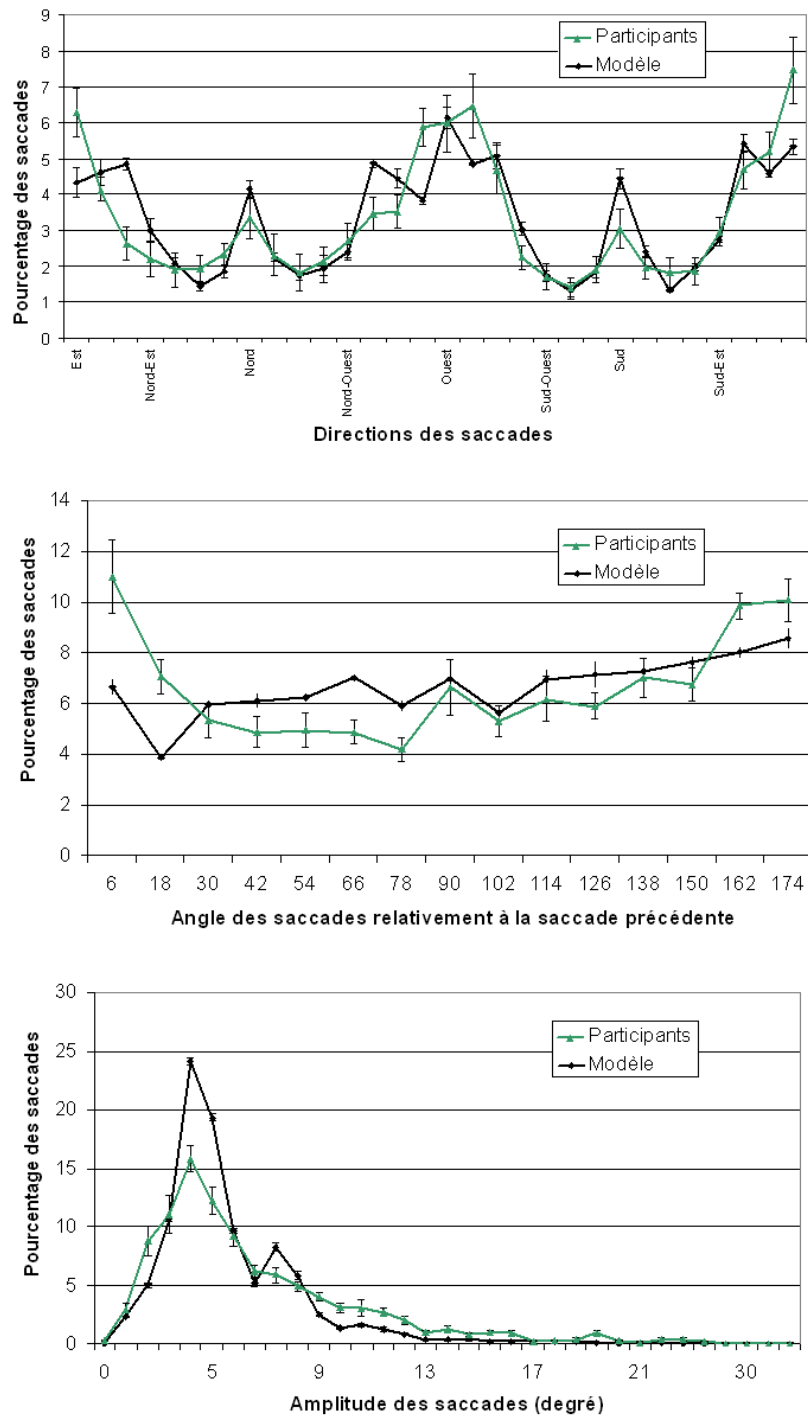


FIGURE 3.21 – Comparaisons entre les distributions empiriques et simulées pour la version *Taille* de l'expérience.

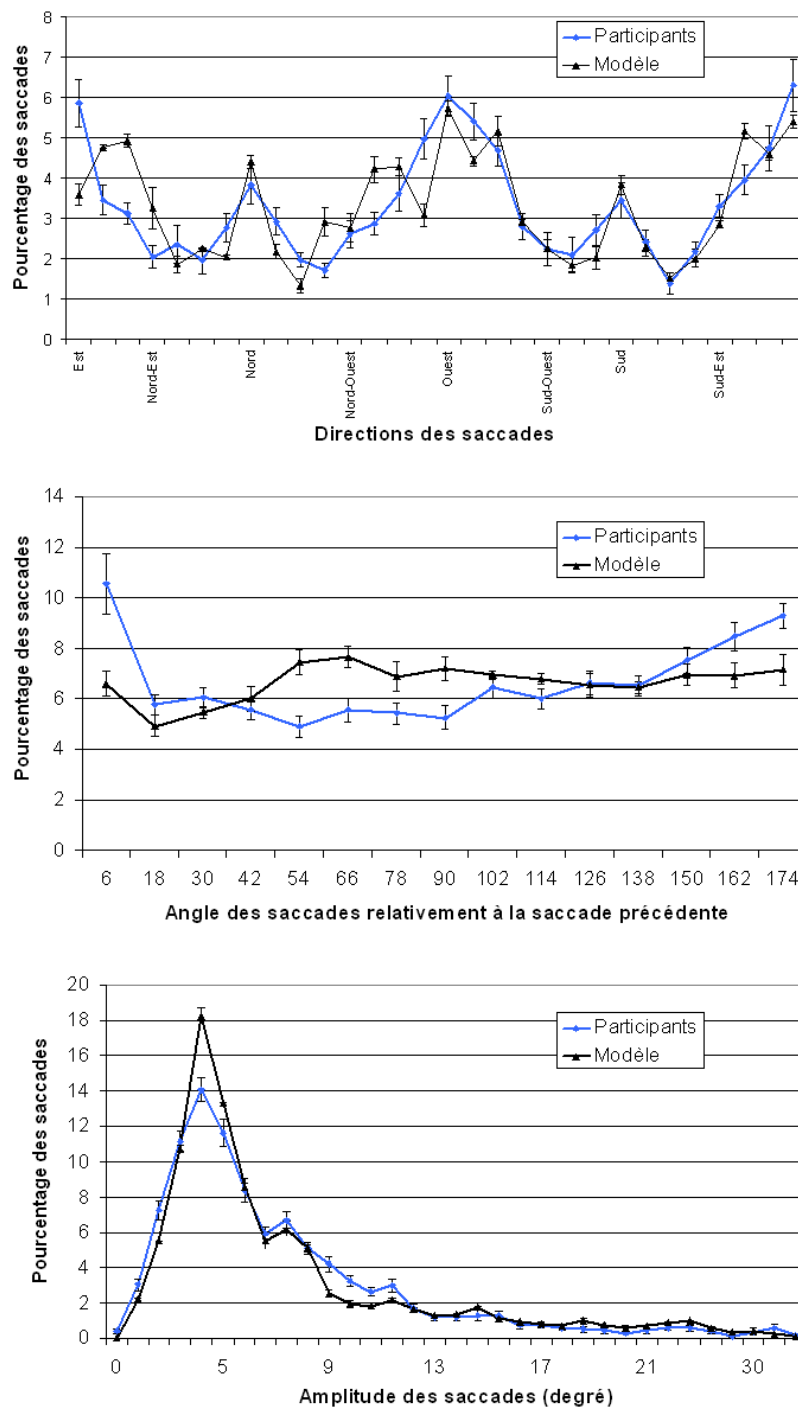


FIGURE 3.22 – Comparaisons entre les distributions empiriques et simulées pour la version *Couleur* de l'expérience.

4

Environnement complexe : matériel expérimental

La première série d'expérimentations que nous venons de voir nous a permis de poser les bases du modèle, mais la tâche que nous cherchons à modéliser et à simuler est la recherche d'information dans un document structuré, idéalement une page Web, même si pour le moment nous n'envisageons pas toute la complexité d'une telle interface. Ce chapitre présente le nouveau matériel expérimental.

Dans cette nouvelle phase de conception du modèle nous allons considérer des éléments plus complexes que des mots, pour lesquels plusieurs fixations sont nécessaires à leur traitement. Les stimuli ne seront plus sous forme de nuages de mots, mais de paragraphes organisés spatialement dans la page. La tâche de recherche d'information est de trouver le paragraphe qui traite le mieux d'un thème donné. Cette seconde série d'expérimentations va nous permettre d'augmenter notre modèle par des mécanismes de lecture et de compréhension de texte, mais aussi des mécanismes de stratégie de parcours car la structure du document est cette fois importante.

Pour l'application de notre modèle à des documents structurés plus complexes, nous avons choisi une tâche de recherche d'information dans un pseudo-journal en ligne. Avec le même matériel expérimental, mais organisé différemment, nous avons défini deux séries d'expérimentations.

4.1 Choix des textes

Le but de cette tâche est de trouver le bloc de texte qui correspond le mieux à l'expression donnée, à laquelle nous ferons référence comme le « thème » de la recherche. Voici quelques exemples de thèmes : *réchauffement climatique, hausse de la bourse, victoire des footballeurs, réforme de l'enseignement...* Étant donné que le but de la tâche est de trouver le bloc de texte qui correspond le mieux au thème, les utilisateurs sont encouragés à analyser tous les blocs sans toutefois sacrifier leur efficacité, car même s'ils ont déjà trouvé un bloc intéressant relativement au thème, le suivant peut être encore plus adéquat.

Nous avons conçu vingt pages contenant à la fois des informations textuelles et des caractéristiques visuelles. Les pages sont composées de sept textes, l'un d'eux étant associé à une page. Les textes sont divisés en trois catégories : deux textes sont sémantiquement très proches du thème et sont donc des cibles potentielles (Association Forte, AF), deux textes ont un rapport sémantique, mais faible, avec ce thème (Association Faible, Afa) et trois textes n'ont pas de lien sémantique avec le thème (Aucune Association, AA). Voici un exemple de sept textes associés au thème *Observation des planètes* :

AF1 - *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.*

AF2 - *L'étude du système solaire est difficile. Écartons 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.*

Afa1 - *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.*

Afa2 - *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.*

AA1 - *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.*

AA2 - *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.*

AA3 - *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.*

Un texte traite des anneaux des planètes géantes, l'autre d'une étude du système solaire : ce sont les textes fortement associés au thème (AF). Un texte concerne les

problèmes de l'antenne d'un engin spatial et un autre d'un logiciel de photographies aériennes : ces deux textes ont une association faible avec le thème (Afa). Les trois autres n'ont pas d'association avec le thème (AA). Ils parlent d'une plainte déposée par une société, de la nageuse Laure Manaudou et de l'après-carrière d'un joueur de basket-ball.

Pour la conception de ces pages, T. Baccino a sélectionné des paragraphes (issus d'articles de diverses sources) fortement, moyennement et faiblement associés à chacun des vingt thèmes choisis. Il les a ensuite modifiés pour constituer des textes de longueur homogène. Chaque paragraphe est constitué d'en moyenne 30,65 mots ($\sigma = 2,78$) formant en moyenne 1,79 phrases ($\sigma = 0,59$). Voici les statistiques de similarité sémantique :

- AF : moyenne de 0,7, max 0,92, min 0,45 ;
- Afa : moyenne de 0,19, max 0,46, min 0,14 ;
- AA : moyenne de 0,01, max 0,13, min -0,05.

Comme nous l'avons vu dans la description de LSA, un seuil classique de 0,2 permet de discriminer les similarités. Les paragraphes AF sont au-dessus de ce seuil, les paragraphes Afa au niveau du seuil, et les paragraphes AA sous le seuil.

Nous avons choisi cent-quarante textes au total répartis dans les vingt thèmes de recherche d'information (un thème par ensemble de sept textes) qui ont ensuite été disposés de manière ordonnée en lignes et en colonnes ou en pêle-mêle dans la page. Tous les textes se trouvent en annexe F page 154.

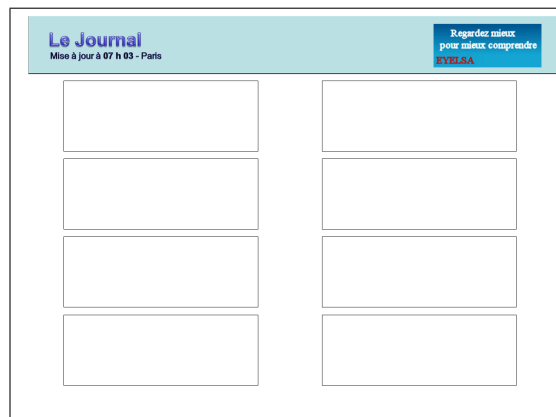
4.2 Organisation en lignes et colonnes

Pour la première utilisation de ce matériel, nous avons placé les paragraphes de manière très structurée, sur une grille de 4×2 : quatre lignes et deux colonnes. Pour avoir un rendu plus réel à notre tâche il a été ajouté un bandeau au-dessus de la page, qui est le même dans toute l'expérience, et indique qu'il s'agit d'un journal. L'organisation générale des blocs est représentée dans la figure 4.1 page suivante

Étant donné que notre modèle est destiné à simuler les conflits que peuvent rencontrer les humains entre les attractions sémantiques et visuelles, nous avons ajouté des informations visuelles à nos pages en plus de l'information sémantique contenue dans les paragraphes.

4.2.1 Information visuelle : photo

Pour être plus proche d'une véritable page Web et avoir une plus grande variabilité sémantique et visuelle, nous avons aussi ajouté une photo sur chaque page. Cette photo est associée à un bloc AF (dans quatre pages sur vingt) ou à un bloc AA, qui se trouve juste en-dessous. Il n'y a aucune photo associée à un bloc Afa ce qui permet d'éviter des situations de renforcement du lien sémantique qui n'aurait pas été quantitativement contrôlé, avec un risque que le lien sémantique entre le thème

FIGURE 4.1 – Organisation en lignes et colonnes (4×2).

et le bloc Afa soit profondément modifié au point de changer sa classe sémantique (vers AF ou AA).

4.2.2 Information visuelle : blocs en couleur

Dans cette première version de cette expérience, les blocs peuvent avoir un fond blanc ou de couleur. Il y a trois couleurs par page, qui ont été choisies de manière à avoir (1) le même contraste entre elles (ces trois couleurs forment un triangle équilatéral dans l'espace CIELab), (2) le même contraste avec le texte écrit en noir et (3) le même contraste avec le fond blanc. Dans chaque page, un bloc AF, un Afa et un AA ont un fond en couleur. Ces trois blocs sont choisis aléatoirement mais de manière uniforme.

4.2.3 Contraintes de placement

Une procédure de randomisation (carré latin) sur les couleurs et les places dans la page a permis de créer six réplifications par thème. Nous disposons ainsi de $6 \times 20 = 120$ pages ce qui correspond à six listes (un participant ne voyant qu'une seule liste). Le placement des paragraphes dans les huit régions ainsi définies a été contrôlé par différentes contraintes de placement. Chaque colonne doit contenir un texte de chaque type (AA, Afa et AF), et les paragraphes en couleur ne doivent pas être l'un au-dessus de l'autre. Il doit donc y avoir une colonne avec deux paragraphes en couleur et une colonne avec un paragraphe en couleur et la photo (exemple de page figure 4.2 page suivante). Pour voir plus en détail le matériel expérimental, les six pages d'un thème (*Croissance de l'économie*) se trouvent en annexe G page 161 et toutes les pages d'une liste (liste 1) figurent en annexe H page 163.

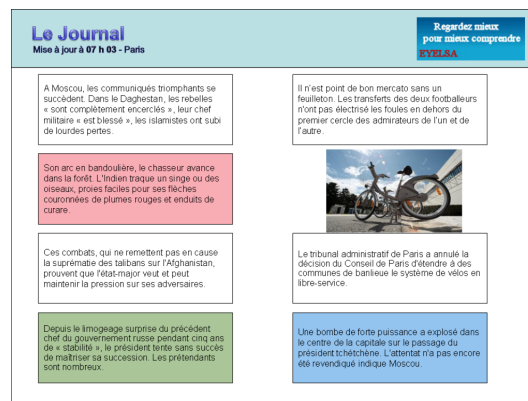


FIGURE 4.2 – Exemple d'organisation en lignes et en colonnes. Objectif de la recherche, trouver le paragraphe qui parle du chef de la Russie.

4.3 Évaluation du matériel

Suite à la première expérimentation avec le matériel organisé en lignes et en colonnes, nous avons analysé les réponses des participants, et notamment leur confiance dans leur choix. En effet, à la fin de chaque recherche ils devaient attribuer un indice de confiance dans leur choix entre 1 et 5. Cette analyse nous a révélé que pour deux pages parmi les vingt la relation entre le thème et les paragraphes AF n'était pas assez proche sémantiquement. La conséquence de ce biais expérimental est une chute dans la confiance dans le choix effectué et une hausse des temps de scrutation (c'est-à-dire le temps pour résoudre la tâche, figure 4.3 page suivante). Les deux thèmes concernés sont notés (bi) et (fu) sur le graphe et correspondent respectivement à *Bombardement en Irak* et *Formation universitaire*. Pour ces deux pages, nous avons aussi une augmentation du nombre de revisites sur les différents paragraphes. Ces deux thèmes seront donc changés pour l'expérience en pêle-mêle.

Dans ces deux pages, les photos étaient associées à des blocs AF. Suite à cette observation, et avant de mettre en place la seconde expérimentation nous avons aussi effectué une vérification du lien sémantique entre la photo et le texte auquel nous l'avions associé afin de savoir si la photo était bien liée au bloc qu'elle illustre.

4.3.1 Test de la photo

Nous avons créé un jeu de cartes en papier contenant chaque paragraphe et chaque photo (8*20 cartes). Nous avons ensuite demandé à vingt personnes, n'ayant pas participé à l'expérimentation, de choisir à partir des huit cartes d'un thème (la photo et les sept textes) les deux textes les plus proches de la photo et de leur attribuer une note entre 1 et 5 indiquant la force de l'association sémantique.

Les résultats montrent que pour les seize thèmes où la photo est attachée à un bloc AA celle-ci est toujours rattachée au bon bloc avec un lien fort (moyenne=4,66), et quasiment toujours éloignée de tous les autres (lien très faible pour le second,

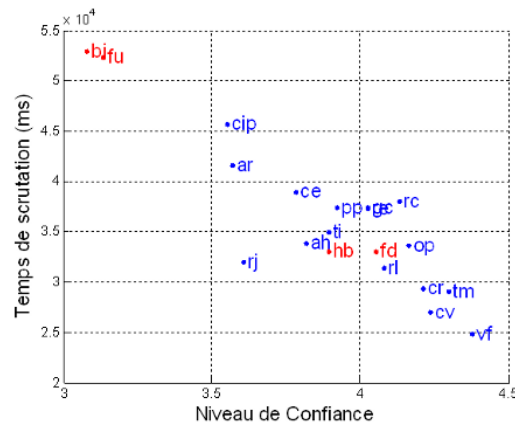


FIGURE 4.3 – Pour chaque thème, indice de confiance par rapport au temps de scrutation, en moyenne sur tous les participants. Les noms en rouge et en gras représentent les pages où la photo est associée à un paragraphe AF.

moyenne=1,89). Pour les quatre thèmes où la photo est associée à un paragraphe AF, les photos sont notées comme proche d'un paragraphe AF en premier (moyenne du premier=4,05), et dans trois cas sur quatre il n'y a pas de distinction possible entre les deux AF (moyenne du second=3,07), ce qui est attendu étant donné que ces deux textes sont proches sémantiquement. Pour le thème *bombardement en Irak* l'association n'est pas clairement exprimée par les participants, et nous changeons donc cette photo. Ce test nous a donc confirmé que les associations que nous avions supposées entre les textes et la photo étaient exactes à une exception près.

4.3.2 Modification du matériel

Suite à ces constats nous avons effectué quelques modifications dans le matériel. Tout d'abord nous avons modifié les thèmes *Bombardement en Irak* et *Formation universitaire* en *Conflit en Irak* et *Formation informatique*. Suite à cette modification les indices de confiance ainsi que les temps de scrutations sont revenus dans la normale dans l'expérience en pêle-mêle.

Les photos ont été modifiées pour deux cartes, *Conflit en Irak*, et *Course des voiliers*. Le nouveau paragraphe auquel une photo a été associée pour *Conflit en Irak* a été un paragraphe AA nous avons donc changé une autre photo pour garder quatre photos associées avec un paragraphe AF, c'est ainsi qu'un paragraphe AF de *Course des voiliers* s'est vu attribué une photo (à la place d'un paragraphe AA).

Avec ces légères modifications, et quelques corrections orthographiques qui nous avaient échappées, nous avons pu construire la seconde partie du matériel expérimental, organisé cette fois en *pêle-mêle*.



FIGURE 4.4 – Copie d’écran du site d’information en ligne du Nouvel Observateur (www.nouvelobs.com).

4.4 Organisation en pêle-mêle

À partir des mêmes textes, nous avons souhaité élargir le test du modèle en complexifiant la répartition des blocs sur la page. En effet, comme nous le verrons par la suite, les trajets inter-blocs des participants sur les pages organisées en lignes et en colonnes sont très systématiques. Afin de ne plus avoir ce biais de structure nous avons donc conçu ce nouveau matériel expérimental, qui est aussi plus réaliste, et se rapprochant de ce que nous pouvons trouver en ligne actuellement, par exemple sur la page du *Nouvel Observateur* (figure 4.4).

Nous avons choisi tout d’abord d’avoir deux formats de cadre contenant les textes. Trois cadres ont un format rectangulaire de 300×219 px, trois ont un format plus allongé de 630×93 px, quant au paragraphe associé à une photo, il est de dimensions 300×342 px et la photo se trouve cette-fois-ci sous le texte auquel elle est attachée (exemple de gabarit figure 4.5 page suivante).

La répartition des blocs AF, Afa et AA a été faite de façon à ne pas avoir les deux blocs AF côte à côte, ni les deux blocs Afa (exemple figure 4.6 page suivante). Avec ces formats, plusieurs dispositions des blocs dans la page sont possibles, nous en avons sélectionné vingt. Pour chaque thème, nous avons choisi quatre dispositions différentes des blocs (un exemple des quatre dispositions pour le thème *Prix du pétrole* se trouve annexe I page 168).

Ce matériel se divise en deux modalités différentes, une *Avec Titre* et une *Sans Titre*. Dans la modalité *Avec Titre* T. Baccino a attribué un titre à chaque paragraphe, comme nous en trouvons couramment dans les pages Web. Ce titre est le nom de la rubrique (Sport, International, Nature, Finance, etc.) se référant au texte, ce n’est pas un titre au sens du résumé de texte. Nous avons donc huit listes : quatre

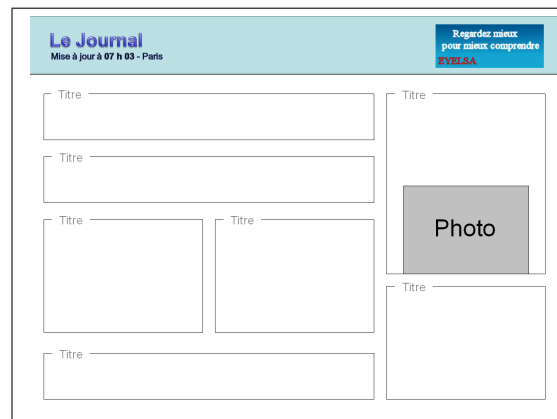
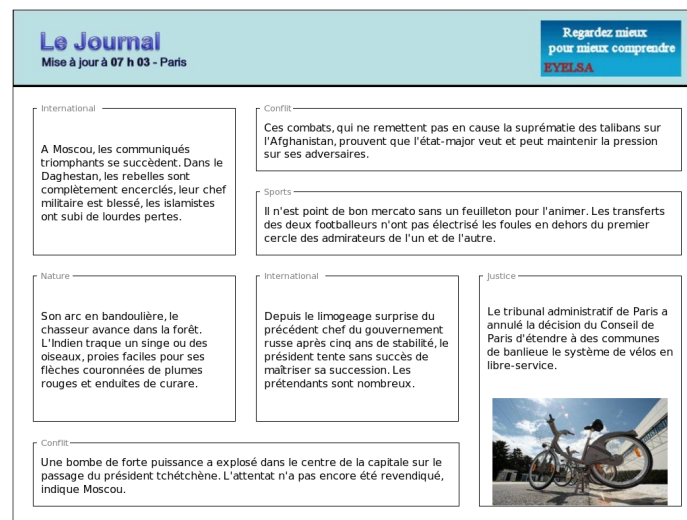


FIGURE 4.5 – Organisation pile-mêle avec titres.

FIGURE 4.6 – Exemple d'organisation en pile-mêle. Objectif de la recherche : trouver le paragraphe associé au thème *Chef de la Russie*.

Sans Titre et quatre *Avec Titre* (liste complète *Sans Titre* en annexe J page 170 et liste complète *Avec Titre* en annexe K page 175). Les chapitres suivants décrivent la modélisation des trajets oculaires sur ce matériel.

5

Environnement complexe : parcours inter-blocs

Pour explorer les différents mécanismes impliqués dans une tâche de recherche d'information dans un environnement complexe, nous avons choisi de faire une scission en deux niveaux. Le premier niveau, qui est l'objet de ce chapitre, concerne la navigation entre les blocs constitutifs du document. Le second niveau, plus local, concerne les fixations à l'intérieur des blocs, et donc dans le cas de paragraphes, la lecture de ces blocs. Cette analyse fera l'objet du chapitre suivant.

5.1 Modélisation du parcours inter-blocs

Étant donné une page et un thème de recherche notre modèle est conçu pour simuler le trajet oculomoteur d'un utilisateur moyen, de manière cognitivement plausible. La production du modèle est donc une séquence de fixations, c'est à dire une suite de coordonnées XY. Nous décrivons ici le premier niveau de la modélisation de la navigation par le parcours inter-blocs pour deux types d'organisation de blocs.

Au niveau du parcours inter-blocs, une observation des trajets enregistrés nous a conduit à différencier deux temps dans le parcours oculaire. Ces deux phases correspondent aux parcours planifié et réactif que nous avons identifiés précédemment. La première phase correspond à une stratégie globale (planifiée) de visite de tous les blocs, avec une recherche d'optimisation du trajet qui respecte le sens d'exploration visuelle des tableaux (parcours de haut en bas et de gauche à droite) (Bouju & Sperandio, 1979) et les contraintes physiologiques (les grandes saccades sont coûteuses).

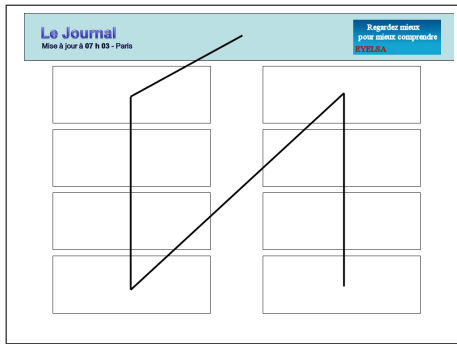


FIGURE 5.1 – Trajet en N inversé généré par le modèle sur une organisation en lignes et en colonnes avec les paramètres $\alpha_H=1,5$ et $\alpha_G=2$.

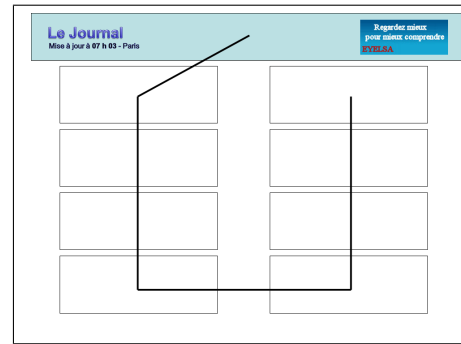


FIGURE 5.2 – Trajet en U généré par le modèle sur une organisation en lignes et en colonnes avec les paramètres $\alpha_H=1$ et $\alpha_G=2$.

Cette planification des premières visites des blocs correspond à la notion de *gist* que nous avons vu dans la section 2.3 page 36. Elle se construit dès l'apparition de la page, avant la première fixation. Une fois tous les blocs visités, une seconde partie du parcours, au niveau local (réactif), concerne les revisites des blocs considérés intéressants. Le modèle reproduit ces stratégies. Depuis le point de fixation initial, il produit ainsi un chemin entre les blocs. Comme pour l'expérience avec les mots, le critère d'arrêt est fixé à partir des données empiriques car le modèle n'a pas de processus lui permettant de prendre cette décision.

5.1.1 Premières visites des blocs

Comme nous l'avons décrit dans le chapitre 2 (page 36) le modèle a des paramètres α_H et α_G qui correspondent respectivement au poids relatif d'un déplacement vers le haut, et au poids relatif d'un déplacement vers la gauche pour lui permettre de caractériser les déplacements entre les blocs.

Plus α_H est grand, plus une transition vers le haut est coûteuse relativement à une transition vers le bas. De même, plus α_G est grand plus une transition vers la gauche est coûteuse par rapport à une transition vers la droite. Si ces poids valent 1, ces transitions sont aussi coûteuses que des transitions vers le bas ou vers la droite. Par exemple, avec $\alpha_H=1,5$ et $\alpha_G=2$, le modèle fait un trajet en N inversé, c'est-à-dire descendra la première colonne de blocs, ira en haut de la seconde puis la descendra (voir figure 5.1). Ce trajet sera généré car il ne contient qu'un seul déplacement vers la gauche (du point de fixation au premier bloc) et qu'une seule transition vers le haut, même si celle-ci est très longue. Si par contre nous changeons les paramètres, et que nous prenons $\alpha_H=1$ et $\alpha_G=2$, le modèle générera un trajet en U, étant donné que les transitions vers le haut ne seront pas pénalisées (voir figure 5.2).

Cet algorithme peut être appliqué à des configurations de blocs plus complexes (type pêle-mêle), comme le montre par exemple la figure 5.3 page suivante.

Chacun de ces processus peut être considéré comme générant une carte de chaleur, avec ici exactement huit zones distinctes. À chaque étape, ces cartes sont additionnées pour former une seule carte globale à partir de laquelle le meilleur bloc va être sélectionné puis fixé. Le poids total P est, pour chaque bloc B_i :

$$P(B_i) = \alpha_{MS} \times P_{MS}(B_i) + \alpha_V \times P_V(B_i) \quad (5.2)$$

Ces deux processus n'ont pas le même rôle. Dans l'expérience précédente, les stimuli étaient des mots isolés, mais les processus impliqués étaient proches. Ce qui était au niveau des mots dans l'expérience précédente est maintenant au niveau des blocs. α_V avait été optimisé à environ 0,35, nous avons donc gardé cette valeur. Pour α_{MS} nous avons pris la somme des paramètres α_M et α_S , c'est-à-dire 0,65. Dans cette nouvelle étude la somme des paramètres est toujours égale à 1.

Comme nous le verrons dans les sections suivantes, cette simple combinaison de stratégies permet déjà de simuler correctement les comportements observés.

5.2 Expérimentations

Pour les deux types d'organisation des blocs, chaque essai commence par une instruction suivie d'une croix de fixation. Cette croix est placée en haut de la page afin d'être loin du texte, dans une zone non informative. Après que les participants aient fixé ce point, la page apparaît. Les participants doivent trouver le meilleur bloc en fonction du thème, sans contrainte de temps, mais en ayant un souci d'efficacité dans la recherche, puis faire apparaître le curseur de la souris en appuyant sur la barre d'espace, et enfin cliquer sur le bloc sélectionné et donner leur confiance dans leur choix. Les participants ont été invités à répondre aussi rapidement que possible, sans sacrifier la précision et il leur a été signifié que plusieurs réponses étaient possibles. La figure 5.4 illustre cette procédure pour le thème *Observation des planètes*, avec les blocs organisés en lignes et en colonnes.

5.2.1 Dispositif

Pour l'expérience avec les blocs en lignes et en colonnes les mouvements oculomoteurs ont été enregistrés avec un oculomètre SR EyeLink II, à une fréquence d'échantillonnage de 500 Hz. Les données n'ont été enregistrées que pour l'œil directeur. L'écran est placé à 50 cm du participant, et entraîne un champ visuel horizontal d'environ 40 degrés. Un repose-menton a été utilisé pour conserver la tête du participant stable. Pour l'expérience pêle-mêle, les mouvements oculaires ont été enregistrés avec un oculomètre SR EyeLink 1000 par T. Baccino, à une fréquence d'échantillonnage de 1000Hz, là encore les données n'ont été enregistrées que pour l'œil directeur. Dans les deux cas les pages font 1024×768 pixels.

L'oculomètre enregistre les positions horizontales et verticales du regard (coordonnées XY) en pixels. De ces coordonnées, les saccades et les fixations sont extraites, ce qui conduit à des trajets comme le montre par exemple la figure 5.5.

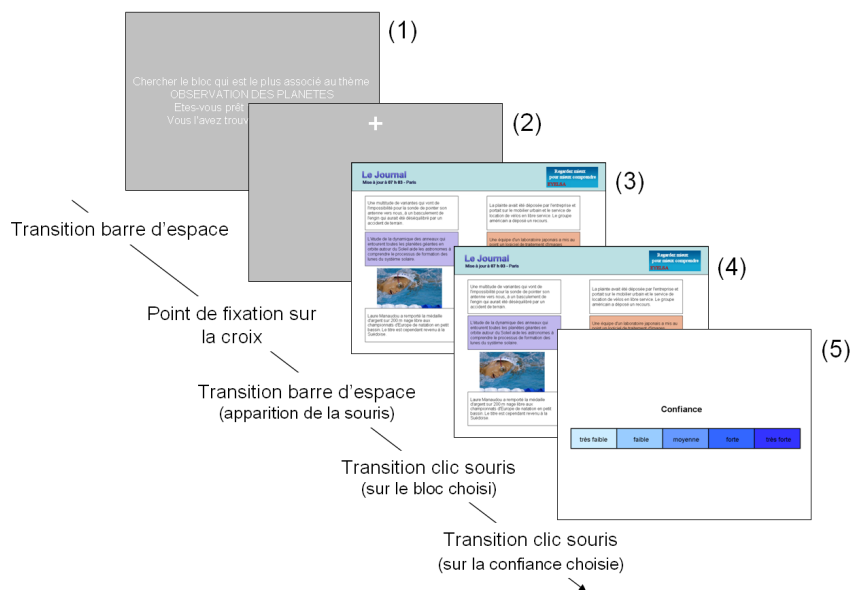


FIGURE 5.4 – Étapes du processus expérimental. Les participants (1) lisent le thème ; (2) fixent le point de fixation ; (3) cherchent l'information dans la page ; (4) donnent leur réponse en cliquant sur le bloc choisi ; (5) donnent leur confiance dans ce choix. Exemple pour le thème *Observation des planètes*.

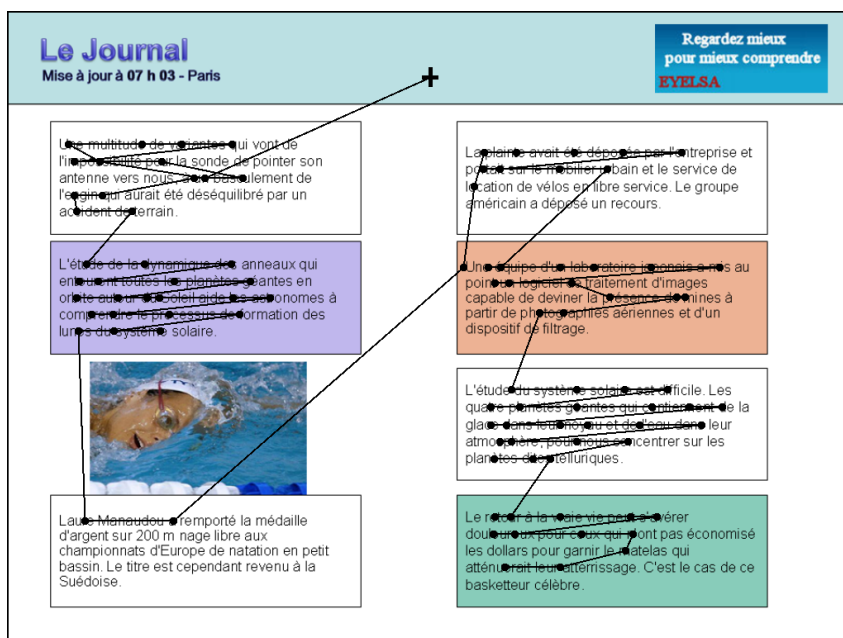


FIGURE 5.5 – Exemple d'un trajet produit par un participant sur le thème *Observation des planètes* dans la version en lignes et en colonnes.

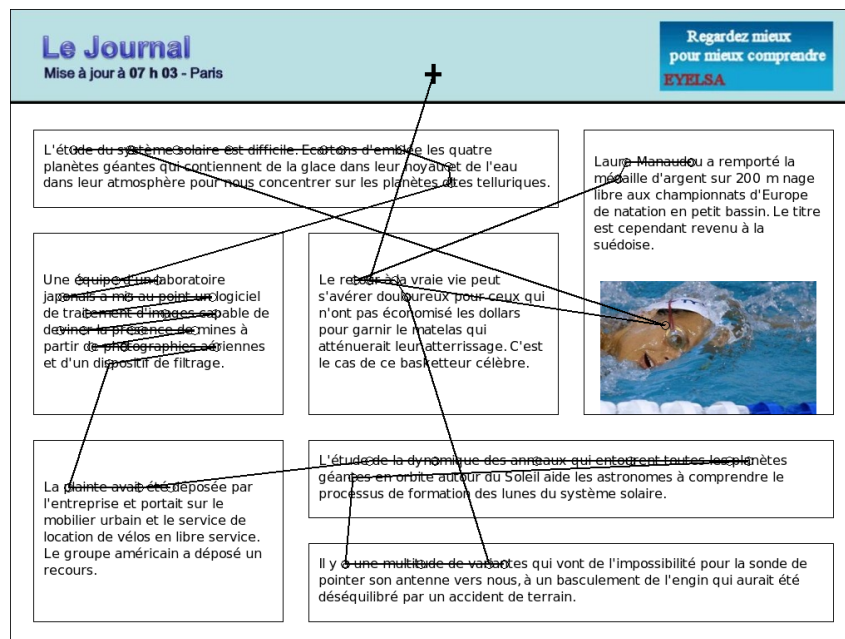


FIGURE 5.6 – Exemple d'un trajet produit par un participant sur le thème *Observation des planètes* dans la version en pêle-mêle.

5.2.2 Participants

Trente-huit participants ont pris part à l'expérience avec les blocs organisés en lignes et en colonnes, et vingt-sept pour celle avec une organisation en pêle-mêle. Tous avaient une vue normale ou corrigée et aucun ne connaissait l'objectif de l'expérimentation.

Comme indiqué précédemment, il existe vingt thèmes différents et six listes. Chacun des trente-huit participants de la première expérimentation a vu une liste contenant tous les thèmes, ce qui conduit à 760 trajets. Pour la seconde expérimentation, les vingt-sept participants ont été répartis dans les deux modalités *Avec Titre* et *Sans Titre*. Chaque participant a vu une liste contenant tous les thèmes. Dans cette expérimentation nous avons dû trier les trajets car certaines données n'étaient pas exploitables. Finalement, 506 trajets ont été gardés (270 pour la modalité *Avec Titre* et 236 pour celle *Sans Titre*).

Notre objectif grâce à ces expérimentations est de tester et valider le modèle. Mais avant de pouvoir comparer le comportement du modèle à celui des participants, nous avons ajusté les paramètres du modèle grâce aux données empiriques.

5.3 Ajustement du modèle

Notre modèle effectue la même tâche que les participants sur les pseudo pages Web. Comme nous l'avons vu précédemment, deux paramètres (α_H et α_G) modulent la stratégie globale du modèle et donc la forme du trajet.

Il n'y a pas de valeurs uniques pour ces paramètres étant donné la variabilité des trajets d'un individu à l'autre, mais il nous fallait déterminer des plages de variations des paramètres correspondant à la variabilité inter-individuelle observée. La méthode employée a été tout d'abord d'observer les comportements empiriques afin d'établir la distribution de ces trajets. En parallèle de ces observations, nous avons déterminé quels étaient les trajets générés par le modèle selon les couples de paramètres α_H et α_G . À partir de ces deux informations, nous avons ajusté les intervalles de α_H et α_G pour obtenir une distribution semblable. Les données issues de l'expérience en lignes et en colonnes ont donc servi à l'apprentissage du modèle. Les données issues de l'expérience en pêle-mêle serviront, elles, au test du modèle.

Nous nous sommes basés sur l'expérience avec les blocs organisés en lignes et en colonnes pour générer tous les trajets possibles issus des combinaisons de α_H et α_G dans l'intervalle raisonnable [0-5] avec un pas de 0,1 et en avons extrait les six trajets les plus fréquents (dont les trajets en U et N inversé). Nous avons ensuite cherché à savoir quelles étaient les stratégies des participants. Pour cela, nous avons classé tous les trajets empiriques selon leur distance avec les trajets prototypiques, en ne prenant que les premières visites sur les blocs. La mesure de distance utilisée est celle de Levenshtein (Navarro, 2001).

La distance de Levenshtein mesure la similarité entre deux chaînes de caractères. Elle correspond au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à l'autre. Un trajet est codé en une chaîne de caractères représentant les blocs numérotés de 1 à 8. Par exemple le trajet en U est codé "12348765" et le trajet en N inversé "12345678". La distance entre ces deux trajets est de 4.

Pour cette répartition des trajets, seuls les trajets dont la distance minimale avec un trajet prototypique était inférieure à 4 ont été pris en compte, les trajets n'ayant que peu de transitions sont donc indéterminés. Les trajets se trouvant à la même distance de deux trajets prototypiques ont été comptés comme appartenant aux deux classes.

À partir de cette répartition, nous avons implémenté un algorithme simple de recherche des intervalles pour α_H et α_G pour lesquels le modèle générait la répartition la plus proche. Ces intervalles sont ainsi [0,7-2,3] pour α_H et [1-2,4] pour α_G . La figure 5.7 montre la répartition des trajets du modèle sur les blocs organisés en lignes et en colonnes dans cet espace à deux dimensions. Chaque couple (α_H, α_G) correspond à un trajet. Parmi les trajets ainsi générés, seuls deux ne faisaient pas partie des six trajets prototypiques. Les deux paramètres α_H et α_G permettent d'une manière simple de bien rendre compte de la variabilité de trajets, comme nous pouvons le voir en comparant les répartitions dans le tableau 5.1 ou sur l'histogramme correspondant (figure 5.8 page suivante).

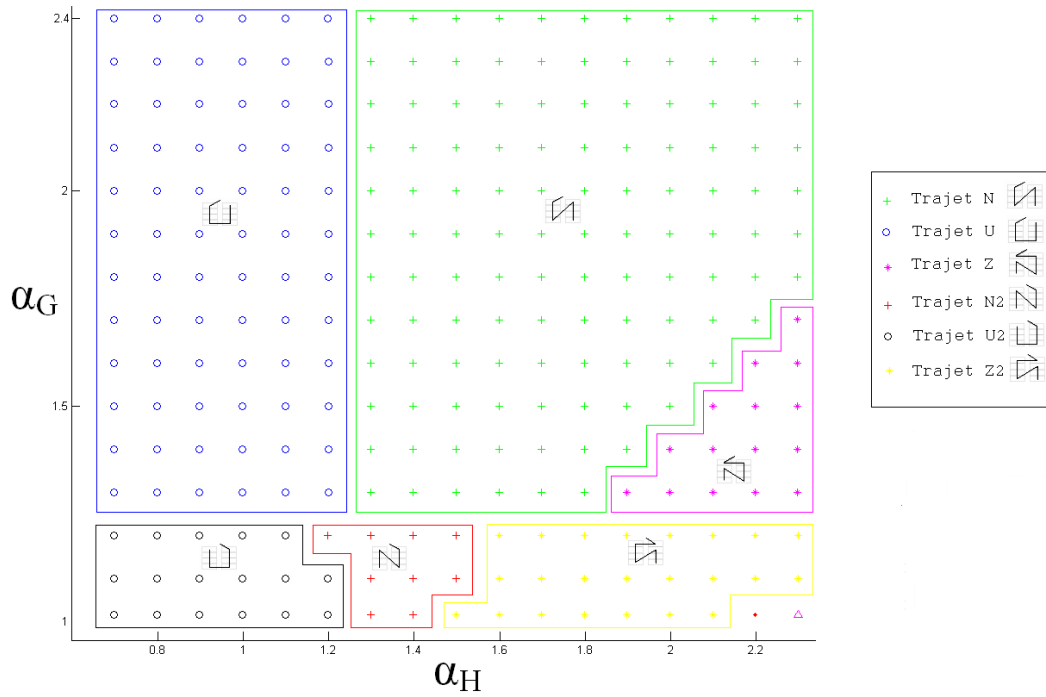


FIGURE 5.7 – Répartition des trajets du modèle selon les deux dimensions.

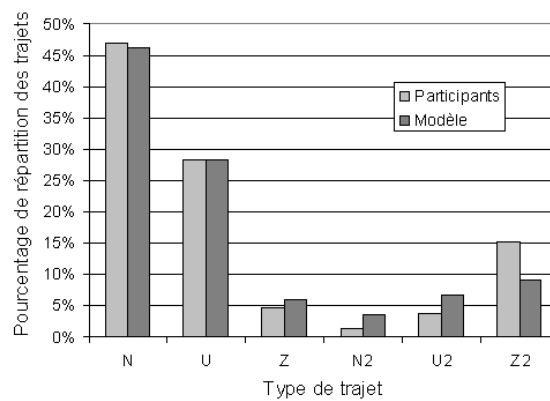


FIGURE 5.8 – Répartition des trajets des participants et du modèle selon le type de trajet.







Trajets Proto	Participants			Modèle		
	Nombre	% Total	% Classés	Nombre	% Total	% Classés
N : 	348	39,32%	46,90%	117	45,88%	46,25%
U : 	209	23,62%	28,17%	72	28,24%	28,46%
Z : 	35	3,95%	4,72%	15	5,88%	5,93%
N2 : 	11	1,24%	1,48%	9	3,53%	3,56%
U2 : 	27	3,05%	3,64%	17	6,67%	6,72%
Z2 : 	112	12,66%	15,09%	23	9,02%	9,09%
Indéterminés	143	16,16%		2	0,78%	
Total	885			256		

TABLE 5.1 – Répartition des trajets de participants et du modèle sur les pages avec les blocs structurés en lignes et en colonnes entre les six trajets prototypiques.

Génération des instances du modèle

Une fois les intervalles de ces paramètres déterminés nous avons tiré aléatoirement trente-six combinaisons de paramètres dans ces intervalles pour simuler trente-six instances du modèle, à comparer avec les trente-huit participants. Nous avons choisi de prendre trente-six instances car nous avons trente-huit participants ; nous avons donc pris le multiple de 6 le plus proche de 38 (car nous avons six listes différentes). Ces trente-six instances ont été réparties dans les six listes, générant 720 trajets (36×20 thèmes).

Pour l'expérience avec les blocs organisés de manière moins structurée (pêle-mêle), nous avons gardé les mêmes intervalles afin de valider la robustesse de notre modèle. Le modèle a tourné sur les huit listes (quatre dans la modalité *Sans Titre* et quatre pour la modalité *Avec Titre*). Trente-deux instances du modèle ont ainsi été simulées (à comparer avec les vingt-sept participants), générant 640 trajets (320 pour la modalité *Sans Titre* et 320 pour la modalité *Avec Titres*).

Pour chaque simulation, nous avons arrêté le modèle, qui n'a pas de condition d'arrêt, quand il avait atteint le nombre moyen de blocs vus par les participants sur la page.

Validation du paramétrage

À l'issue de cette première expérimentation, nous avons cherché à savoir si les participants et le modèle ont une même stratégie de parcours des blocs avec d'autres indicateurs que ceux qui nous ont permis de paramétrer le modèle. Nous avons ainsi observé que la première colonne (celle de gauche) est presque toujours parcourue de haut en bas, à la fois par les participants (79,07%) et le modèle (96,25%). Celle de droite est parcourue vers le haut pour 27,14% des participants, et 39,30% des instances du modèle, et vers le bas pour 62,92% des participants et 56,38% pour le

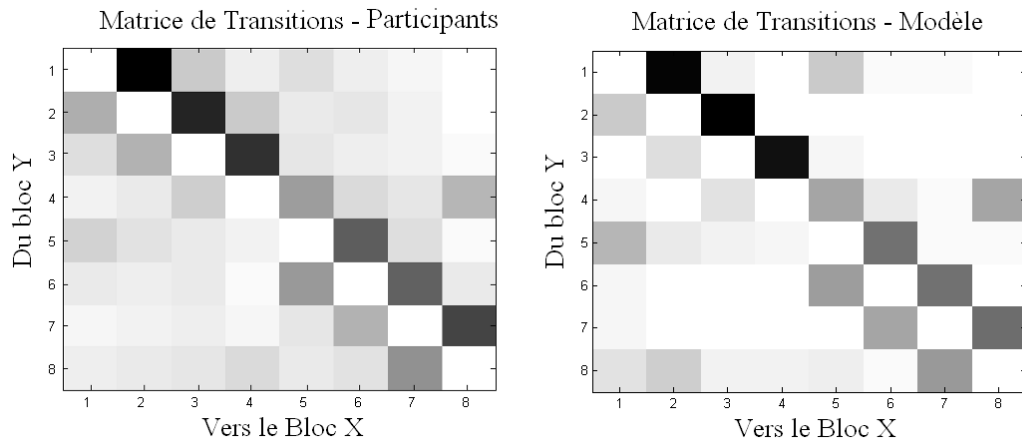


FIGURE 5.9 – Matrice de transitions entre les blocs pour les participants et le modèle. Les blocs sont numérotés de 1 à 4 pour la première colonne et de 5 à 8 pour la seconde.

modèle. Ces caractéristiques n’ont pas été directement optimisées par le paramétrage du modèle, elles en sont néanmoins très liées. Nous nous attendions ainsi à cette correspondance. Nous pouvons toutefois noter que le modèle a une tendance plus forte que les participants à descendre la colonne de gauche, ainsi qu’à monter celle de droite.

Des matrices de transitions (figure 5.9) entre chacun des huit blocs nous montrent aussi les similitudes entre les trajets des participants et ceux du modèle. La couleur de la case (i,j) dénote le nombre de transitions du bloc i au bloc j . Plus la couleur est foncée, plus le nombre de transitions est important. Les blocs sont numérotés de 1 à 4 pour la première colonne et de 5 à 8 pour la seconde.

Ces matrices montrent qualitativement qu’avec un simple algorithme de planification de chemin et deux paramètres indiquant la tendance à aller vers le bas et vers la droite, il est possible de reproduire assez finement la stratégie globale des parcours dans cette tâche de recherche d’information.

5.4 Comparaison du modèle avec les données empiriques

5.4.1 Transitions entre les blocs

Organisation pêle-mêle

Il est important de noter que dans cette phase de validation du modèle les paramètres n’ont pas été modifiés, et sont donc ceux ajustés dans les expérimentations précédentes. Les trajets générés par le modèle sur ce matériel expérimental vont donc nous permettre de tester sa robustesse, sans apprentissage supplémentaire des paramètres à partir des données.

Étant donné l’organisation de la page et la variabilité du matériel expérimental, une matrice de transition générale ne peut pas être générée. Nous aurions pu géné-

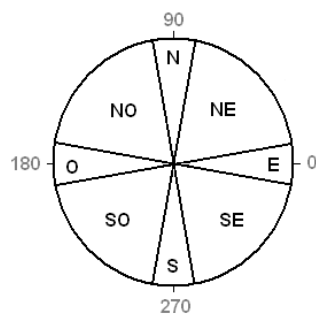


FIGURE 5.10 – Définitions des directions de transitions entre les blocs.

rer une matrice par trajet, ou par organisation, mais étant donné le nombre d'organisations différentes ces matrices n'auraient pas pu nous apporter d'information significative sur l'émergence de trajets principaux prototypiques. Dans l'organisation précédente du matériel expérimental nous avons pu numéroter les blocs de 1 à 8 sur les deux colonnes, et ainsi comptabiliser le nombre de transitions du bloc x au bloc y pour tous les trajets. Dans cette deuxième organisation du matériel, les blocs ne sont plus alignés, et il y a vingt organisations différentes des blocs dans la page. Pour néanmoins comparer les trajets, nous avons mis en place un autre indice : les directions des transitions. Chaque transition entre deux blocs a été qualifiée par une direction (N, NE, E, SE, S, SO, O ou NO), et ce pour tous les trajets des participants et du modèle. Chaque transition a été déterminée comme allant du centre du bloc quitté au centre du bloc suivant, afin d'avoir une validation du comportement du modèle au niveau du bloc. Une transition est Nord-Est si sa direction est dans l'intervalle $[10 - 80]$, une transition est Nord si sa direction se trouve dans l'intervalle $[80 - 100]$, une transition est Nord-Ouest si sa direction est dans l'intervalle $[100 - 170]$, une transition est Ouest si sa direction est dans l'intervalle $[170 - 190]$, etc (voir figure 5.10).

Ces distributions de direction, effectuées sur les premières visites des blocs représentent donc les transitions entre blocs, et nous permettent d'analyser la propension du modèle à reproduire les comportements observés (voir figure 5.11 page suivante). Sur cette figure, nous voyons que le modèle a tendance à effectuer plus de transitions vers le sud que les participants. Nous pensons qu'une partie de ce biais peut-être expliqué par la position de la photo qui se situe toujours sous le bloc auquel elle est associée. En effet, le modèle ne traite pas la photo mais il effectue une fixation par trajet au centre de celle-ci. Ainsi lorsqu'il a fini de traiter le bloc associé à la photo, il fait une fixation sur la photo, donc va vers le sud, puis va sur le bloc suivant. Dans chaque parcours du modèle il y a donc au moins une fixation vers le sud qui est comptabilisée alors que pour les humains la tendance à fixer systématiquement la photo n'est pas aussi forte. Par contre cette figure nous montre que les paramètres α_H et α_G définis dans l'expérience précédente, permettent au modèle d'avoir un comportement proche de celui des participants, avec peu de transitions vers la gauche, et peu de transitions vers le haut.

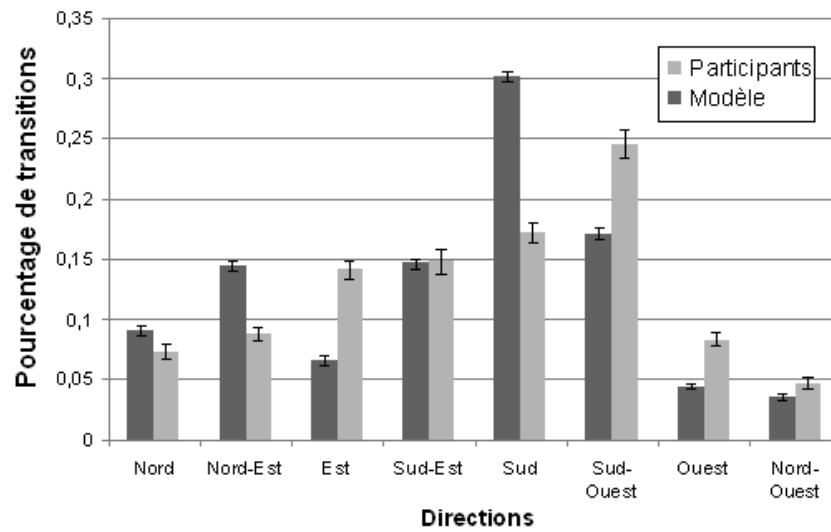


FIGURE 5.11 – Graphe directionnel des transitions entre les blocs pour les participants et le modèle.

5.4.2 Revisites des blocs

Organisation en lignes et en colonnes

Le deuxième indice de comparaison entre les humains et le modèle concerne les revisites. Nous avons cherché à savoir si le modèle retourne sur les mêmes blocs que les participants. Lorsque les participants retournent sur des blocs qu'ils ont déjà vus auparavant, nous observons que ce sont ceux qui sont sémantiquement plus proches du thème qui sont le plus revisités. Le modèle a un comportement similaire grâce à la stratégie locale de parcours des blocs. La figure 5.12 montre les taux de revisites. Ces taux correspondent au nombre de revisites sur un type de bloc par rapport au nombre total de revisites. Ils sont nettement plus élevés pour les blocs AF (Association Forte) que pour les autres blocs, et en particulier les AA (Aucune Association) (test de Student, tous les $p < 0,01$) pour le modèle et les participants.

Organisation pêle-mêle

Comme pour l'expérience précédente, les taux de revisites selon le type de bloc nous permettent une bonne comparaison entre les données expérimentales et simulées. Les blocs les plus revisités sont les blocs AF, fortement associés à l'objectif de la recherche et ceux qui le sont le moins sont les blocs sans aucune association avec le thème (voir figure 5.13 page ci-contre). Nous pouvons voir dans cette figure que les taux de revisites ne varient pas, ou très peu, selon la modalité utilisée (*Avec Titre* ou *Sans Titre*), pour les participants comme pour le modèle.

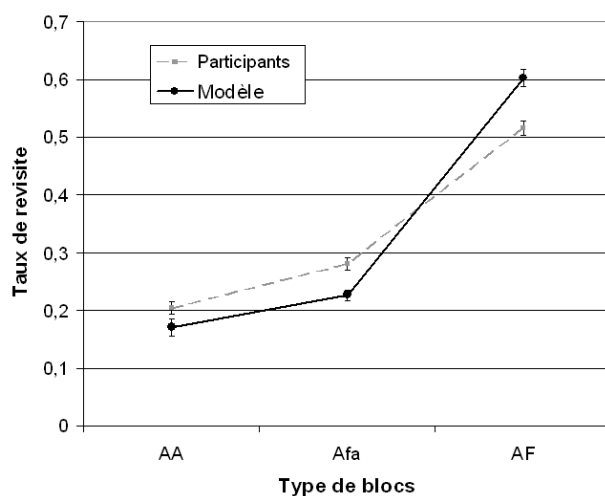


FIGURE 5.12 – Taux de blocs revisités pour les participants et le modèle selon le type de bloc (Association Forte, association faible, Aucune Association) pour l’expérience avec l’organisation en lignes et en colonnes.

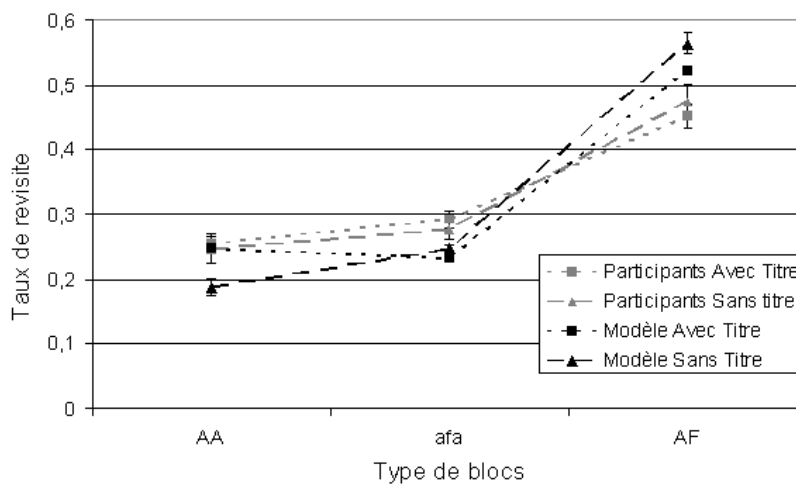


FIGURE 5.13 – Taux de blocs revisités pour les participants et le modèle selon le type de bloc (Association Forte, Association Faible, Aucune Association) pour l’expérience avec l’organisation pêle-mêle.

5.5 Discussion

Dans un environnement plus complexe qu'une simple constellation de mots sur un écran, lorsque les mots sont organisés en phrases, en paragraphes, et que ceux-ci forment différents blocs dans une page, des stratégies de parcours des blocs se mettent en place. Suite à une étude de ces comportements, nous avons pu les modéliser en se basant sur des heuristiques et des mécanismes simples de parcours de documents.

À ce niveau de modélisation, entre les blocs, le modèle reproduit assez fidèlement le comportement des participants au niveau du premier passage sur les blocs dans différentes modalités. Nos choix de modélisation, utilisant simplement deux paramètres, et un algorithme de construction de chemin ont en effet permis, dans une certaine mesure, de reproduire les données empiriques. La disposition des blocs en *pêle-mêle* est en effet un moyen plus rigoureux pour nous de valider le modèle car les paramètres utilisés ici ont été fixés préalablement par l'expérimentation précédente. Nous voyons donc que les tendances qu'avaient les participants à parcourir la page de haut en bas et de droite à gauche se retrouvent dans la seconde expérimentation.

En ce qui concerne les revisites, le modèle reproduit la forte préférence qu'ont les participants à retourner vers les blocs pertinents pour la recherche. Ce mécanisme prend en compte à la fois les informations visuelles et sémantiques contenus dans les blocs pour choisir celui ou ceux à revisiter. Cette double prise en compte d'information est modulée par le phénomène d'oubli simulant la mémoire du parcours. Cette seconde phase du parcours entre les blocs dépend donc en partie du traitement effectué à l'intérieur des blocs. Dans cette phase de revisite, l'ajout d'un titre n'a pas d'effet, pour les participants, comme pour le modèle. Nous verrons par contre l'effet du titre au niveau du parcours intra-bloc.

Notre modèle a néanmoins ses limites. Le traitement de la photo est pour le moment trop sommaire. L'information contenue dans la photo aide les participants à traiter le paragraphe qui y est associé, en particulier lorsque les blocs, et donc la photo, sont éloignés sémantiquement du thème de la recherche. Le modèle devrait donc lui aussi en tenir compte. De plus nous voyons en observant les trajets oculomoteurs expérimentaux que la photo n'est pas forcément fixée, alors que, par construction, elle est fixée systématiquement par le modèle. Au niveau visuel, l'intégration dans le processus visuel d'un modèle de saillance visuelle plus sophistiqué permettrait aussi de mieux prendre en compte la photo, où tout autre élément visuel présent sur la page. Mais il faudrait alors adjoindre des processus de plus haut niveau de reconnaissance de formes.

Dans le prochain chapitre nous allons voir quel est le comportement des participants à l'intérieur des blocs, ainsi que leur modélisation et simulation.

6

Environnement complexe : parcours intra-bloc

Dans le chapitre précédent, nous avons vu quelles étaient les stratégies de parcours de la page au niveau des blocs, pour les participants et le modèle. Ce dernier chapitre traite du parcours intra-bloc dans des documents structurés, c'est-à-dire du comportement oculaire au niveau des mots. Nous n'allons pas revenir ici sur la méthodologie utilisée pour la mise en place des expérimentations et les caractéristiques des différents paragraphes, qui ont été traités dans le chapitre 4. Nous verrons en revanche les résultats observés chez les participants à ce niveau plus fin et la modélisation de ce comportement que nous avons déjà évoqué dans le chapitre 2 page 31. Nous finirons par des comparaisons entre les données empiriques et simulées.

Pour modéliser le comportement oculaire humain sur une tâche impliquant de la lecture, les modèles de lecture nous rappellent que les questions à se poser sont *où ?* et *quand ?* c'est-à-dire où et quand déplaçons-nous notre attention. Dans nos travaux, nous nous sommes focalisés sur la question *où* en laissant de côté la question *quand*. Cette question reste cependant majeure dans les améliorations à apporter à notre modèle étant donné que la simulation des durées de fixations, qui représentent le traitement cognitif par mot, permettra une meilleure compréhension des différents comportements observés, et en particulier concernant les prises de décision.

Dans ce chapitre, nous traiterons avec une attention particulière l'utilisation des titres. Dans l'expérimentation où les paragraphes sont organisés en pêle-mêle, nous avons deux conditions dont une où les paragraphes sont associés à un titre. Ce titre ne représente pas une synthèse du paragraphe, mais il représente la rubrique, la

catégorie des informations qui se trouvent dans le paragraphe, par exemple *Sport*, *International*, *Nature*, *Santé*, etc. Des auteurs en ergonomie des IHM ont étudié l'utilisation de titres comme résumé du paragraphe et se sont intéressés, par exemple, aux différentes fonctions que le titre peut prendre (titre thématique, fonctionnel, générique, relationnel), ou encore à la validité cognitive du titre (Eyrolle, Virbel, & Lemarié, 2008). Dans nos travaux, nous utiliserons le titre comme le nom de la rubrique du paragraphe, en considérant qu'il est effectivement bien associé au paragraphe auquel il est rattaché et qu'il doit être traité de manière singulière lors de sa lecture.

6.1 Résultats expérimentaux

6.1.1 Lecture vs exploration

Les comportements observés sont très variables, mais pour une grande majorité les parcours intra-bloc commencent au début du paragraphe et le suivent ensuite en séquence sur les différents mots. Nous observons néanmoins des fixations qui semblent « isolées » à l'intérieur d'un paragraphe. Pour évaluer le lien sémantique du paragraphe avec le thème, le participant peut mettre en place une stratégie où il « explore » le paragraphe par quelques mots qui lui suffisent, sans passer par une lecture plus systématique. Nous observons aussi des combinaisons de fixations dites « exploratoires » avec de la lecture. Aussi il nous est apparu important de pouvoir discriminer, dans les données expérimentales, les fixations de type *Exploratoire*, et celles de type *Lecture*, qui correspondent à deux stratégies bien distinctes. Cette classification s'est faite à partir de la localisation de la fixation et des précédentes.

Les fixations sont prises par fenêtre glissante de trois et sont classées en *Lecture* si elles sont dans un schéma de lecture, c'est-à-dire soit toutes les trois alignées, soit deux sur une ligne et une sur la ligne suivante ou précédente (voir schémas figure 6.1 page ci-contre), avec des saccades de faible amplitude (distance en $x < 150$ pixels et distance en $y < 25$ pixels). Ces distances représentent en horizontal une distance d'environ trois mots, et en vertical la distance entre le milieu de la ligne courante et le bas de la ligne inférieure, soit un peu plus qu'un interligne (voir figure 6.2 page suivante). Les fixations qui ne sont pas en *Lecture* sont de type *Exploratoire*. Les deux premières fixations de chaque paragraphe ont un traitement particulier. Si la première fixation est proche du premier mot du paragraphe elle est classée en *Lecture*. Si la seconde fixation est alignée à la première, selon les mêmes critères de distance, elle est elle aussi classée en *Lecture*. Un trajet oculaire est ainsi une succession de fixations dont la majorité, dans nos données empiriques, sont en *Lecture*. Sur l'exemple, figure 6.3 page ci-contre, les fixations en vert (clair) sont les fixations en *Lecture* et les fixations en rouge (foncé) sont celles *Exploratoires*.

À partir de cette classification nous avons cherché à savoir plus précisément quelles étaient les différentes stratégies à l'intérieur des blocs et si des profils d'utilisateurs émergeaient. Nous avons donc analysé la répartition des fixations pour chaque

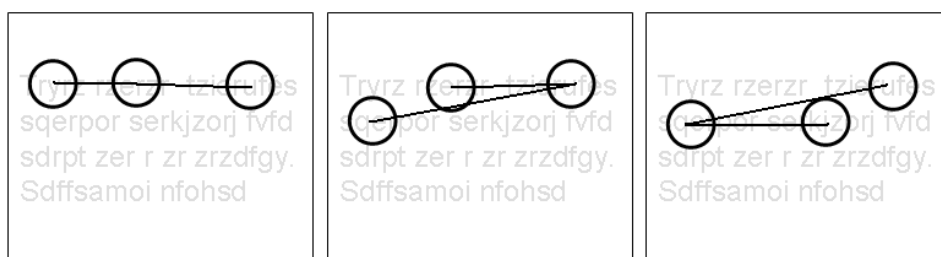


FIGURE 6.1 – Différentes configurations prises en compte pour qualifier les fixations en *Lecture*.

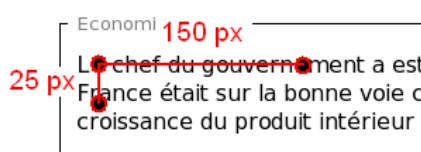


FIGURE 6.2 – Distances maximales en horizontal et vertical pour qualifier les fixations en *Lecture*.

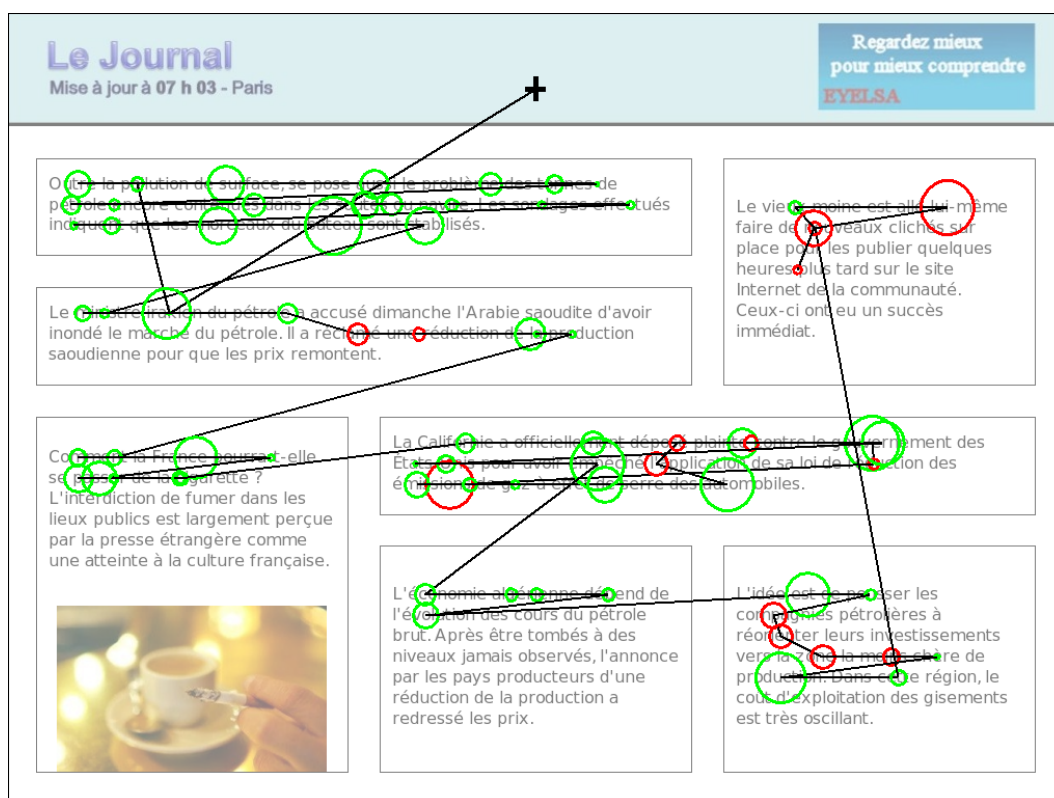


FIGURE 6.3 – Exemple de trajet oculomoteur dont les fixations sont colorées selon leur classe, *Lecture* en vert (clair) et *Exploratoire* en rouge (foncé). La taille des cercles correspond à la durée des fixations.

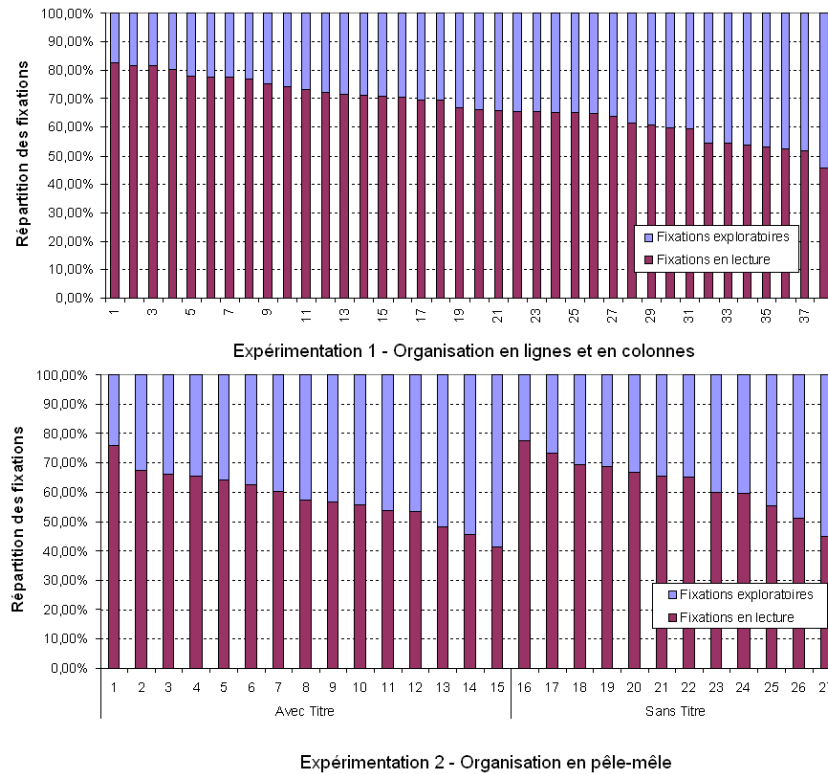


FIGURE 6.4 – Répartition pour les deux expérimentations des fixations selon leur classe (*Exploratoires* ou *Lecture*) par participant. Pour l’organisation en pêle-mêle, les deux modalités, *Avec Titre* et *Sans Titre* sont représentées.

participant (voir figure 6.4). Sur cette figure nous pouvons voir pour les deux expérimentation qu’il n’y a pas de profils d’utilisateur particulier, mais plutôt un continuum de profils et que le nombre de fixations en *Lecture* est majoritaire pour la quasi totalité des participants. Cette observation nous a conduit à faire le choix de ne modéliser que les fixations en *Lecture* (du moins dans la version actuelle du modèle). Nous allons maintenant voir dans la section suivante le cas des paragraphes *Avec Titre*.

6.1.2 Traitement du titre

Lorsque les paragraphes sont accompagnés d’un titre, les comportements des utilisateurs sont particuliers. Afin de modéliser cette nouvelle prise d’information nous avons cherché à savoir quelle était l’utilisation du titre. Les analyses ont été réalisées lors du premier passage sur les paragraphes.

Lors de recherche et de navigation sur le Web, il est très fréquent que les paragraphes, dans les journaux en ligne par exemple, soient accompagnés d’un titre. Dans notre étude, avant de commencer la modélisation, nous avons essayé de corroborer l’hypothèse que cette association titre/texte était connue en vérifiant qu’il n’y avait pas de phénomène d’apprentissage du lien entre le titre et le paragraphe et surtout

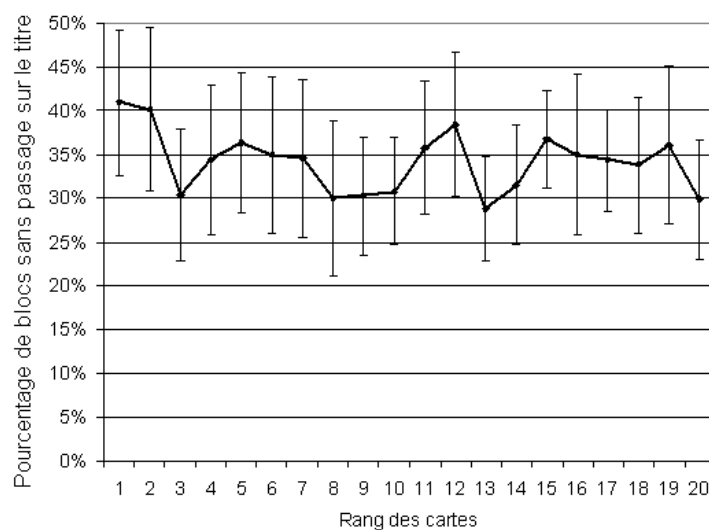


FIGURE 6.5 – Proportion des paragraphes lus sans aucune fixation sur le titre (lors du premier passage), selon l'ordre d'apparition des cartes.

de l'utilité du titre.

Chaque fixation a été indexée selon sa localisation, titre, texte ou photo. Puis nous avons calculé la proportion de paragraphes dont le titre n'était jamais fixé, et ce, selon le rang d'apparition de la carte. Nous pouvons considérer qu'il n'y a pas d'effet du rang de la carte sur cette variable ($F(19) = 0,36$, $p = 0,99$), même si lors des deux premières pages traitées le taux de paragraphes dont le titre n'est pas fixé semble plus important que par la suite (voir sur figure 6.5). Par contre, nous pouvons observer un effet significatif de la variabilité inter-individuelle ($F(14) = 14,22$, $p < 0,01$). Nous avons donc une grande variabilité inter-individuelle mais pas d'effet d'apprentissage.

Quand et comment le titre est-il pris en compte par les utilisateurs? Pour répondre à ces questions, nous avons séparé en quatre classes les paragraphes afin d'avoir une meilleure connaissance des comportements en vue de la modélisation. Ces classes sont les suivantes :

- Texte fixé en premier, puis aucune visite sur le titre. Seul le texte a donc été vu lors de la première visite.
- Texte fixé en premier, puis le titre. En général dans ce cas de figure, le titre est fixé après une ou deux fixations sur le texte.
- Titre fixé en premier puis le texte. Le titre incite à poursuivre la lecture du bloc en allant sur le texte.
- Titre fixé en premier puis une transition vers un autre bloc. Le titre incite à aller vers un autre bloc et à ne pas traiter le texte du bloc.

En regardant les comportements individuels nous observons que la répartition des blocs visités dans ces quatre classes est très variable (voir figure 6.6 page suivante).

En vue de la modélisation, nous avons cherché à caractériser ces phénomènes. Ainsi nous montrons que l'amplitude des saccades arrivant sur un titre est plus

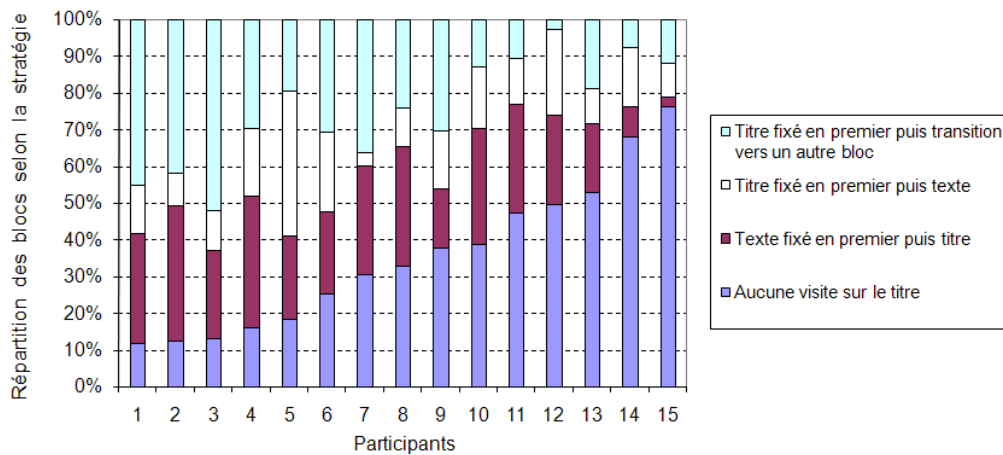


FIGURE 6.6 – Répartition des paragraphes selon l’utilisation du titre (lors du premier passage) pour les différents participants.

courte (262 pixels, $\sigma = 153,48$) que pour celles arrivant sur le texte (327 pixels, $\sigma = 161,91$) ($F(1) = 66,74, p < 0,01$). Notre hypothèse est donc que lors de la programmation de la saccade le titre était visé, étant donné son importance, mais que l’exécution de la saccade n’a pas aboutie à l’endroit prévu. Cet effet est connu sous le nom d’erreur saccadique ou de bruit saccadique. Des études montrent que dans des contextes visuels très pauvres (où les cibles à viser sont des points lumineux), il n’y a pas de relation directe entre l’augmentation de l’amplitude de la saccade et la précision de la position spatiale de la fixation. En revanche, dans un contexte visuel plus riche, comme c’est le cas ici, l’augmentation de l’excentricité dégrade l’apparence visuelle de la cible et peut alors provoquer une augmentation de l’erreur de position de la fixation (Pomplun & Ritter, 1999). Par contre une fois arrivée sur le texte, la localisation de la fixation suivante ne dépend pas de l’amplitude de la saccade, et une fois sur deux le titre est fixé par les participants lors de cette seconde fixation.

6.2 Modélisation

À partir des observations empiriques, que ce soit pour les paragraphes alignés en lignes et en colonnes, ou ceux en *pêle-mêle* (avec ou sans titre) nous avons modélisé un comportement global du modèle. Le modèle effectue une lecture en séquence des paragraphes (mot à mot) et la similarité sémantique entre le texte lu et le thème de la recherche lui permet de prendre la décision de continuer à lire ou de quitter le bloc, comme nous l’avons vu brièvement (section 2.6.2 page 58).

6.2.1 Analyse morpho-syntaxique

Nous savons que la catégorie lexicale des mots joue un rôle dans la lecture. Afin de simuler les fixations sur les mots, nous prenons en compte cette information. Pour

DET : ART	Les
NOM	enfants
VER : pres	jouent
PRP : det	au
NOM	foot

TABLE 6.1 – Résultats de l’analyse morpho-syntaxique de TreeTagger pour la phrase *Les enfants jouent au foot*.

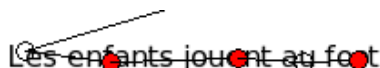


FIGURE 6.7 – Fixations simulées par le modèle sur la phrase *Les enfants jouent au foot*. Les cercles remplis correspondent aux fixations sur les mots traités.

repérer la catégorie lexicale des mots nous avons utilisé le logiciel TreeTagger, développé par Schmid (1994). Cet outil permet un étiquetage automatique des mots. Il attribue à chaque mot une catégorie grammaticale et peut aussi lemmatiser les mots (les mettre sous forme canonique) mais nous ne nous servons pas de cette fonctionnalité dans notre modèle. À partir d’un corpus d’apprentissage étiqueté manuellement, TreeTagger apprend les règles d’étiquetage en appliquant un algorithme d’arbre de décision. Le tableau 6.1 montre un exemple de sortie du programme TreeTagger à partir du texte suivant : *Les enfants jouent au foot*.

À partir de cette analyse, le modèle va identifier et négliger les mots outils, car ces mots sont généralement sous-fixés lors de tâche de lecture chez les humains (King & Kutas, 1995 ; Long, 2000). Ces mots sont les déterminants, les pronoms, les conjonctions, les prépositions, etc. Le modèle considère donc dans son traitement les mots porteurs de sens. Nous qualifierons ces mots de *lexicaux*. Dans notre exemple, seul les mots *enfants*, *jouent* et *foot* seront traités et les mots *les* et *au* seront ignorés (voir les fixations simulées par le modèle figure 6.7).

Les mots lexicaux (dont la catégorie lexicale est informative, nom, verbe, adjectif, etc) sont donc sélectionnés, et la similarité entre ces mots et le thème de la recherche est ensuite calculée par LSA.

6.2.2 Similarité entre le texte et le thème

Une fois le bloc choisi, au niveau du parcours global, le modèle s’y dirige. S’il n’y a pas de titre, le premier mot lexical va être fixé et il va être comparé au thème de la recherche. Si le cosinus entre ces deux vecteurs est supérieur à un seuil donné, cela signifie que la proximité entre le thème et le texte est si forte qu’une lecture plus complète n’est pas nécessaire. Si, au contraire, le cosinus est inférieur à un autre seuil donné, cela signifie que la similarité entre le thème et le texte est si faible qu’il n’est pas nécessaire de poursuivre la lecture. Ces deux seuils, décrits précédemment,

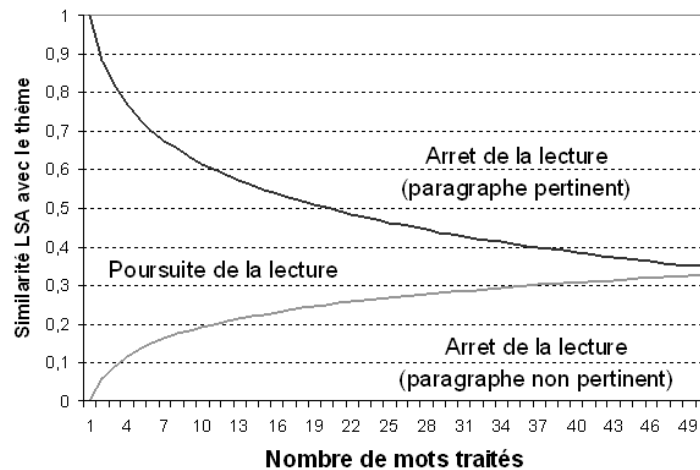


FIGURE 6.8 – Seuils de similarité en fonction du nombre de mots traités.

correspondent aux courbes figure 6.8. Dans ces deux cas la lecture est interrompue et le modèle passe au bloc suivant. Si par contre le cosinus se situe entre ces deux seuils, le modèle continue la lecture et intègre le prochain mot lexical. Une fois sur le mot suivant, le modèle calcule le vecteur contenant les deux mots lexicaux lus et le compare au thème, puis, selon le cosinus entre les deux vecteurs, une nouvelle décision est prise : quitter le bloc ou continuer la lecture. Le modèle continue ainsi sa lecture en prenant à chaque nouveau mot lexical traité une décision, jusqu'à ce qu'il quitte le bloc ou qu'il en fasse une lecture complète. Les seuils sont variables en fonction du nombre de mots traités. Plus il y a de mots plus la confiance dans la similarité est grande et plus il est donc facile de prendre la décision de quitter le bloc.

6.2.3 Prise en compte du titre

Modélisation des fixations sur le titre

Dans le cas des paragraphes accompagnés d'un titre, celui-ci a un statut particulier et le modèle le traite de manière différenciée. D'après les observations des données oculomotrices, l'amplitude de la saccade entrant sur le bloc est déterminante. Comme nous pouvons le voir sur la figure 6.9, la proportion de saccades arrivant sur le titre est plus importante lorsque l'amplitude de la saccade est faible. Cette courbe est fortement dépendante de la distribution des saccades de transition entre blocs (c'est-à-dire les saccades arrivant sur un bloc, figure 6.10). Cette distribution nous permet également de noter ici un biais du matériel expérimental. En effet, étant donnée la disposition des blocs, il y a moins de saccades d'amplitude d'environ 200 pixels (environ 8 degrés d'angle visuel) que de saccades de 100 pixels et de 300 pixels. Ceci est dû au fait que dans une stratégie de déplacement de titre en titre les distances sont fixes et ne sont pas d'environ 200 pixels, mais plutôt de 100 ou 300 pixels (voir figure 6.11 page 122). Le pic vers le bas de la figure 6.9 correspond à un artefact du

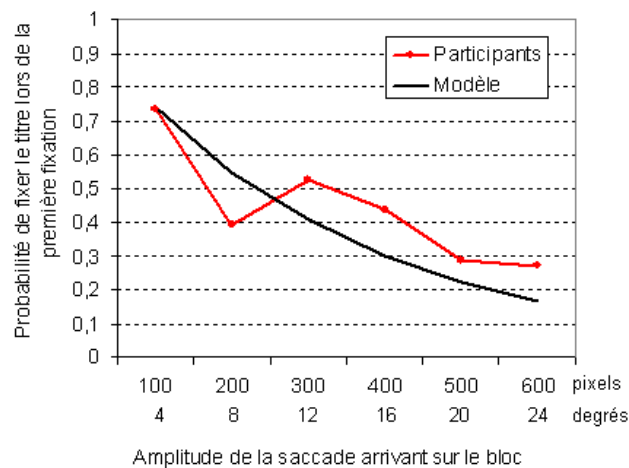


FIGURE 6.9 – Probabilité d’aller sur le titre lors de la première fixation sur le bloc selon l’amplitude de la saccade entrant sur ce bloc.

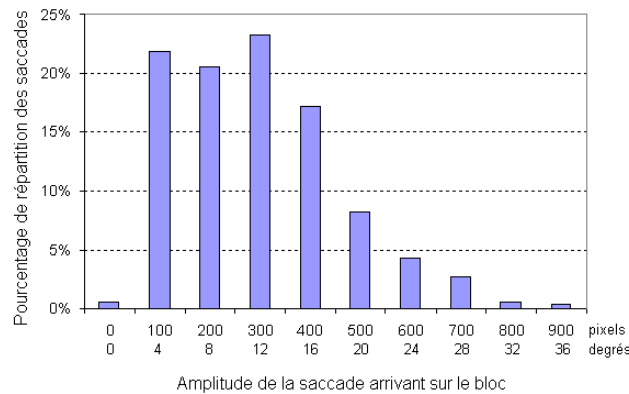


FIGURE 6.10 – Répartition des saccades entrant sur le bloc selon leur amplitude.

matériel que nous n’essaierons pas de modéliser.

Nous avons modélisé globalement le rôle de l’amplitude de la saccade d’une manière simple, en utilisant une loi exponentielle. Une fois le bloc choisi au niveau du parcours inter-blocs, cette loi détermine la probabilité qu’a le modèle d’atterrir sur le titre dépendamment de l’amplitude de la saccade ($AmplitudeS$). Si l’amplitude de la saccade est importante cette probabilité sera plus faible, mais si l’amplitude de la saccade est plus courte la probabilité d’atteindre le titre sera plus grande. La loi représentant ce phénomène est la suivante :

$$e^{(-\gamma \times AmplitudeS)} \quad (6.1)$$

avec $\gamma = 0,3$.

Cette courbe, basée sur les données expérimentales (figure 6.9) ne prend pas en compte les amplitudes de saccades au-delà de 600 pixels (24 degrés angulaires) car des saccades de si grandes amplitudes sont très rares. Il en va de même pour les

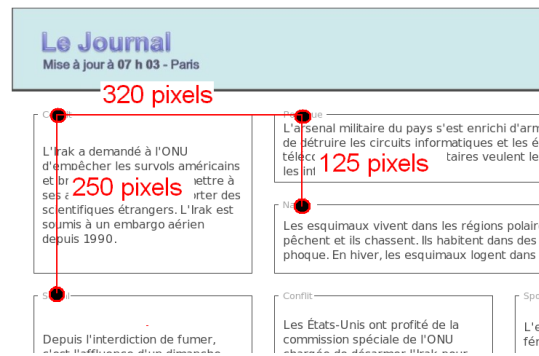


FIGURE 6.11 – Distance en pixels entre les titres des blocs.

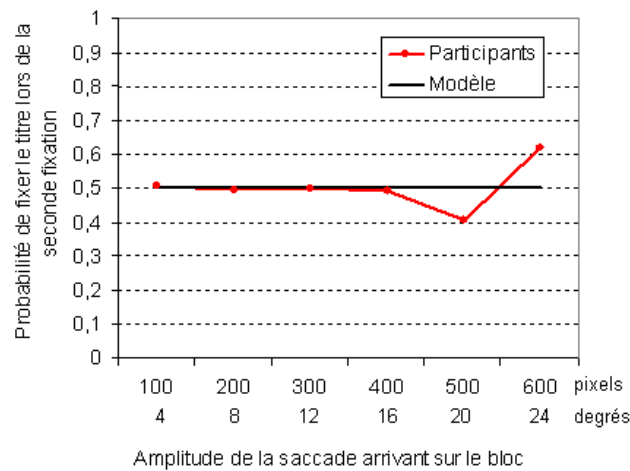


FIGURE 6.12 – Probabilité d’aller sur le titre lors de la seconde fixation sur le bloc selon l’amplitude de la saccade entrant sur ce bloc, si le titre n’a pas encore été fixé.

saccades de très faibles amplitudes inférieures à 50 pixels (2 degrés angulaires) (voir la répartition des saccades arrivant sur un bloc selon leur amplitude figure 6.10).

Si le titre n’est pas fixé lors de l’arrivée sur le bloc, une seconde probabilité est utilisée pour déterminer si la seconde fixation sera sur le titre ou non. À partir des données expérimentales, nous avons observé que cette probabilité ne dépend pas de l’amplitude de la saccade, comme le montre la figure 6.12. Nous avons donc fixé cette probabilité à 0,5. Ainsi lors de la seconde fixation, le titre sera fixé une fois sur deux s’il ne l’a pas été lors de la première fixation. Enfin si le titre n’est toujours pas fixé, il ne le sera plus par la suite. En effet, en observant le comportement des participants, nous voyons que les fixations sur le titre après plus de deux fixations sur le texte sont plus rares (8% des cas).

Lecture du titre

Nous venons de décrire la modélisation de la fixation d’un titre, nous allons maintenant décrire la prise en compte du titre comme élément d’information. Un

titre a une signification particulière lors d'une recherche d'information. En effet, si celui-ci est en lien avec le thème de la recherche, l'intérêt du paragraphe dont il fait référence augmente. Au contraire, si le titre n'est pas du tout relié au thème de la recherche, la lecture du paragraphe devient inutile. Par exemple si le thème est *Faiblesse du dollar*, un paragraphe dont le titre est *Sports* ne paraît pas du tout pertinent, par contre un paragraphe dont le titre est *Économie* paraît intéressant.

Le modèle prend donc cette similarité en compte en calculant le cosinus LSA entre le thème et le titre lorsque celui est fixé. Si ce cosinus est en-dessous d'un certain seuil, le paragraphe n'est pas lu. Le seuil est fixé à 0,2. Cette même valeur a également été utilisée dans les chapitres précédents car il est communément admis qu'à partir de ce seuil les similarités entre termes sont jugées comme pertinentes. Ainsi, seuls les paragraphes vraiment éloignés du thème ne sont pas lus.

LSA est connu pour avoir des performances moindres dans le cas de comparaisons entre deux documents courts (deux ou trois mots). Dans notre cas, la similarité entre le thème (deux mots) et le titre (un mot) est moins bien évaluée par LSA. Ainsi la décision de ne pas lire le bloc si la similarité est inférieure à 0,2 peut paraître brutale et entachée d'erreurs. Nous avons tout de même laissé ce seuil à 0,2, ne sachant pas comment « proprement » résoudre ce problème.

Si la similarité entre le titre et le texte était supérieure à ce seuil, celui-ci est pris en compte pour la comparaison entre les mots traités et le thème. Afin de prendre en compte son importance particulière, son poids est doublé par rapport aux autres mots traités, comme nous le verrons dans l'exemple de trajet page suivante.

Lorsque le modèle quitte un bloc la similarité entre ce bloc et le thème est mémorisée pour les revisites (comme mentionné dans le chapitre précédent). Si un bloc a été abandonné en raison de sa pertinence pour la recherche, sa similarité avec le thème sert de base au mécanisme de mémoire. Il en est de même si le texte est lu dans son intégralité. Par contre si le bloc est abandonné car non pertinent, sa valeur pour le mécanisme de mémoire est initialisée à 0.

Validation des paramètres

Afin de valider l'ajustement des probabilités, nous avons observé la répartition des données empiriques et simulées dans les quatre classes de comportements vis à vis du titre. Nous rappelons que ces classes sont : aucune visite sur le titre ; texte fixé en premier puis fixation sur le titre ; titre fixé en premier puis fixation sur le texte ; titre fixé en premier puis transition vers un autre bloc. Nous obtenons une répartition du modèle proche de celle des données expérimentales à un niveau global, même si les différences pourraient être moindres (voir figure 6.13 page suivante).

Nous allons maintenant décrire un exemple de trajet, avant de passer aux comparaisons entre les trajets simulés par ce modèle et les trajets empiriques.

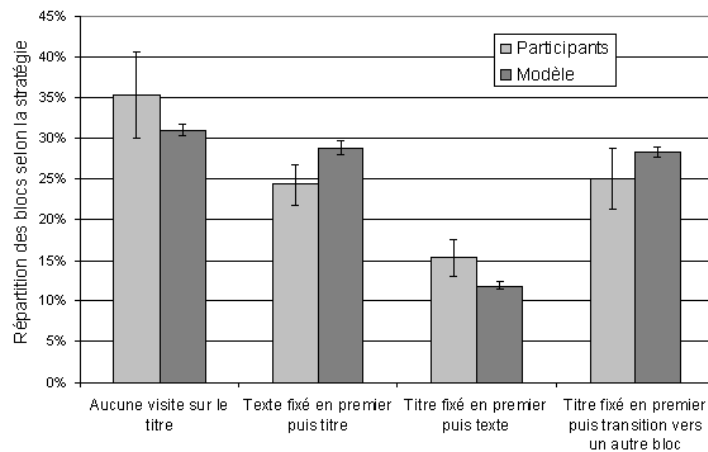


FIGURE 6.13 – Répartition globale des paragraphes selon l’utilisation du titre pour les participants et le modèle.

6.2.4 Exemple de trajet du modèle

Dans ce paragraphe nous allons décrire en détail un trajet effectué par le modèle. Celui-ci est représenté sur la figure 6.14 page ci-contre. La tâche est de trouver le paragraphe le plus proche du thème *Observation des planètes*, qui est rappelé en haut de la figure.

Les paramètres sont $\alpha_H=1$ et $\alpha_G=1$. Le modèle planifie tout d’abord son trajet de bloc en bloc, et choisit de fixer le bloc en haut à gauche. La saccade pour atteindre le titre depuis la croix de fixation initiale est de 436 pixels. Après tirage aléatoire, le titre n’est pas fixé. Le premier mot lexical, *plainte*, est fixé. Sa similarité avec le thème est de -0,06. Le modèle continue sa lecture car il ne quitte pas le bloc sans avoir traité au moins deux mots. Après un nouveau tirage aléatoire, le modèle choisit de ne pas fixer le titre et de poursuivre sur le texte. Les mots *avait* et *été* sont successivement ajoutés au mot *plainte* pour former un vecteur à comparer avec le thème. La similarité est de -0,01. Le modèle quitte donc le bloc.

Un nouveau trajet inter-blocs est généré et le bloc suivant est choisi. Il s’agit du bloc juste en-dessous. L’amplitude de la saccade pour l’atteindre est de 135 pixels et, lors du tirage aléatoire pour ce second bloc, le titre est fixé. La similarité entre le titre *Astronomie* et le thème *Observation des planètes* est de 0,61. Le modèle va donc commencer la lecture du paragraphe puisque cette valeur est suffisamment élevée. Dans les comparaisons des similarités, le titre comptera double dans le vecteur des mots. Par exemple, lorsque la fixation est sur le mot *anneaux* la comparaison se fait entre le vecteur des mots *astronomie astronomie étude dynamique anneaux* et le vecteur du thème *Observation planètes*. Dans ce cas, la lecture se poursuit jusqu’au mot *système*. La similarité a atteint 0,58 et étant donné le nombre de mots traités, celle-ci dépasse le seuil. Le bloc est donc abandonné et sa similarité avec le bloc étant importante, le modèle en garde la trace en mémoire pour une potentielle revisite.

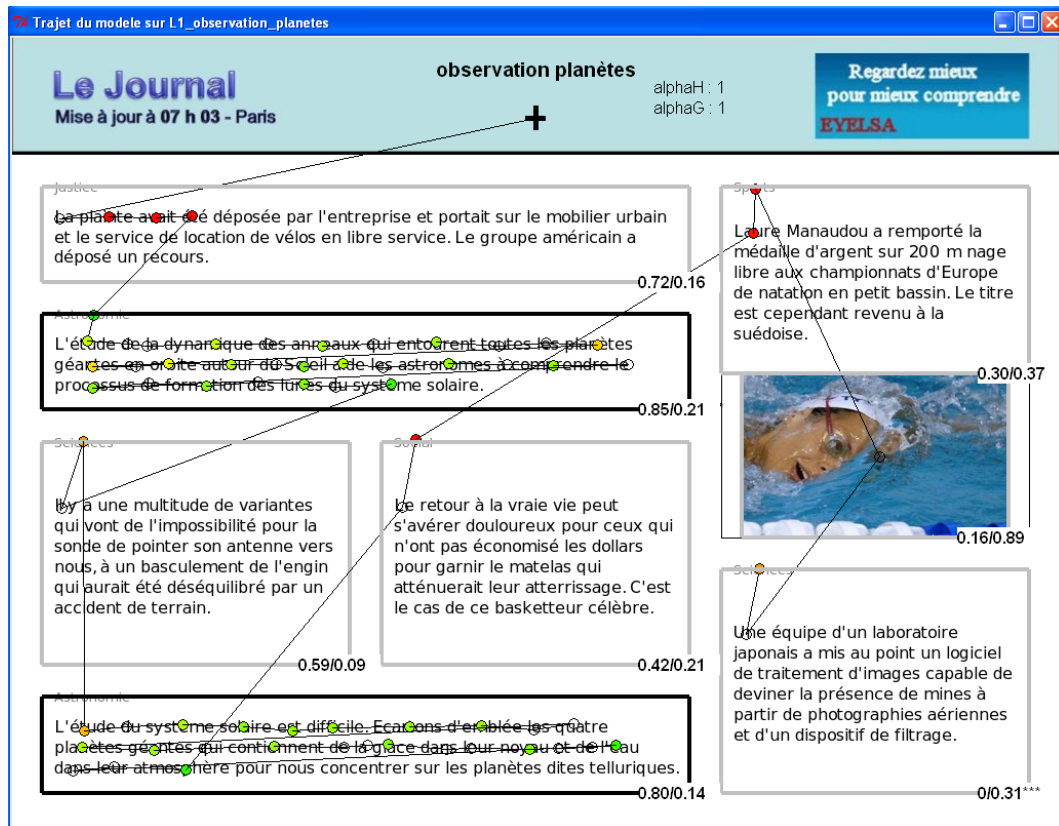


FIGURE 6.14 – Exemple de trajet du modèle sur une carte de l'expérimentation 2 avec titre. Le thème de la recherche est *Observation des planètes*. Dans certains cas le titre est fixé, et selon la similarité entre les mots traités et le thème, les blocs sont abandonnés car non pertinents pour la recherche (entourés en gris), ou abandonnés mais mémorisés pour des revisites (entourés en noir) ou lus dans leur globalité si aucune décision n'a pu être prise (non représenté ici). Les valeurs en bas à droite de chaque bloc représentent les valeurs dans les cartes visuelle et sémantico-mnésique, utilisées lors des revisites.

Un nouveau trajet inter-blocs est généré et le bloc suivant est choisi, là encore il s'agit du bloc juste en-dessous. Cette fois-ci, la saccade de 305 pixels n'atterrit pas sur le titre. Celui-ci est néanmoins fixé après une première fixation sur le texte. La similarité entre le titre *Sciences* et le thème *Observation des planètes* de 0,19 ne dépasse pas le seuil de 0,2 et le paragraphe ne sera donc pas lu. Nous voyons ici la tendance un peu trop brutale du modèle à ne pas lire des paragraphes qui pourraient pourtant être pertinents pour la recherche.

Le modèle passe donc au bloc suivant, le titre (après une saccade de 250 pixels) n'est pas fixé, et la lecture du texte se poursuit jusqu'à *atmosphère*. Le bloc est jugé intéressant pour la suite. Le modèle continue son parcours jusqu'à avoir visité tous les blocs. Il utilisera ensuite les valeurs sémantiques mémorisées pour les revisites.

Sur cet exemple nous pouvons aussi noter les valeurs mnésiques et visuelles des blocs pour les revisites (en bas à droite de chaque bloc). La valeur à gauche est celle de la mémoire, et à de droite celle de la carte visuelle. Pour une première revisite, ce serait donc le deuxième bloc en haut à gauche (celui intéressant) qui serait choisi. Cet exemple nous a permis de bien comprendre les conséquences des choix de modélisation, nous allons maintenant décrire comment comparer les données empiriques et simulées, et quels sont les résultats de ces comparaisons.

6.3 Comparaison du modèle avec les données empiriques

Afin de comparer les données expérimentales et celles produites par le modèle sur les stratégies de lecture intra-bloc, nous avons sélectionné différentes variables. La première est le taux de lecture dans le paragraphe, et les suivantes concernent l'utilisation du titre. Nous avons observé les trajets expérimentaux, et étant donné la grande variabilité inter-individuelle, nous avons utilisé des variables robustes descriptives des comportements oculomoteurs en lecture, pour établir la comparaison entre les données expérimentales et simulées.

6.3.1 Taux de lecture

Le taux de lecture, calculé pour chaque bloc, est le nombre de fixations effectuées par rapport au nombre de mots présents, lors de la première visite sur les blocs. Cet indice va nous permettre d'évaluer l'intérêt porté au paragraphe. Le modèle traite les mots en séquence : ce taux de lecture représente donc le ratio entre le nombre de mots traités et le nombre total de mots du paragraphe. Pour les participants, nous ne connaissons pas le nombre de mots lus, nous utilisons donc le nombre de fixations, qui nous a semblé un bon indicateur du temps de traitement et de l'attention portée au paragraphe, même si les fixations ne sont pas en séquence. Le taux de lecture pour les participants correspond donc au ratio du nombre de fixations sur le nombre total de mots du paragraphe. Ceci n'est pas un taux à proprement parlé car sa valeur peut-être supérieure à 1, si par exemple, le participant parcourt plusieurs fois le paragraphe avant de le quitter.

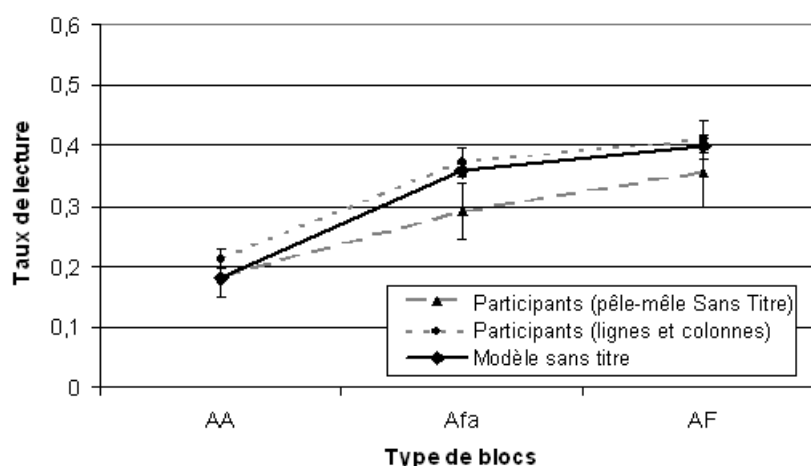


FIGURE 6.15 – Comparaison des taux de lecture (nombre de fixations/nombre de mots dans le paragraphe) lors de la première visite entre les données empiriques et simulées pour les organisations en lignes et en colonnes et pêle-mêle sans titre.

Lorsque les paragraphes ne sont pas accompagnés d'un titre (à la fois lorsque le matériel est structuré en lignes et en colonnes, et lorsqu'il est en pêle-mêle) nous obtenons des taux de lecture pour les participants et le modèle relativement proches (voir figure 6.15). Les taux de lecture empiriques et simulés sont en effet plus importants lorsque le bloc est associé au thème de la recherche (AF et Afa), alors qu'ils sont plus faibles lorsque le bloc est éloigné du thème (AA). Les seuils décrits précédemment ont donc bien permis, de manière relativement simple, de rendre compte de cette stratégie de lecture qui consiste à quitter un bloc lorsque celui-ci n'est pas pertinent.

Lorsque nous nous intéressons à la modalité *Avec Titre*, les résultats montrent aussi que le modèle simule bien le comportement des participants pour les blocs non pertinents pour la recherche (voir figure 6.16 page suivante). Cependant pour les blocs intéressants (Association Forte) la différence entre le modèle et les participants est importante. En observant le comportement du modèle, nous pouvons voir que lorsque le modèle fixe un titre, il a ensuite tendance à fixer plus de mots pour pouvoir prendre sa décision. Dans l'exemple de trajet décrit précédemment la différence entre les deux blocs intéressants (entourés en noir) est très claire : dans un cas, le titre *Astronomie* est fixé et le modèle lit quasiment tout le bloc avant de le quitter, alors que dans l'autre, le titre n'est pas fixé et le modèle quitte le bloc à peine la troisième ligne entamée. Pour les participants, en revanche, le titre permet de survoler le bloc, et moins de fixations sont nécessaires au traitement du paragraphe, même dans le cas de paragraphes intéressants. Afin de pallier à ce problème dans la modélisation, une piste envisageable serait de moduler les seuils en fonction du titre au lieu de simplement doubler le poids du titre dans le traitement du paragraphe, ou encore de simuler un comportement plus exploratoire du paragraphe.

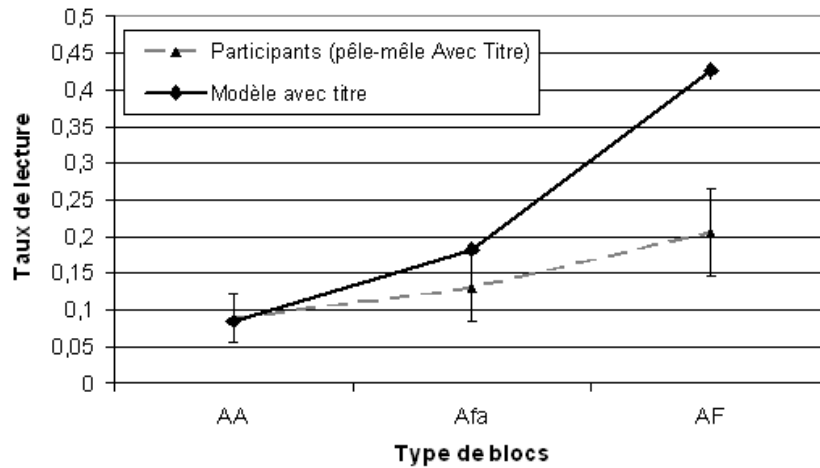


FIGURE 6.16 – Comparaison des taux de lecture (nombre de fixations/nombre de mots dans le paragraphe) lors de la première visite entre les données empiriques et simulées pour l'expérimentation 2 dans les 2 modalités.

Pour les paragraphes avec une association faible (Afa), deux effets se compensent. Lorsque le titre est fixé les blocs ne sont pas forcément lus (si la similarité entre le titre et le thème est inférieure à 0,2) ; par contre si le titre n'est pas fixé, ou que la similarité est supérieure au seuil de 0,2, le paragraphe est ensuite très fréquemment lu dans la globalité.

Les seuils qui ont permis une bonne simulation des données empiriques dans le cas de paragraphes *Sans Titre* restent donc adaptés aux paragraphes plus ou moins éloignés du thème de la recherche, mais ne le sont plus dans le cas de paragraphes proches du thème. La robustesse de notre modèle atteint donc ici une limite dans sa capacité à intégrer de nouveaux types d'éléments dans la page.

6.3.2 Utilisation du titre

Nous avons vu que les paramètres utilisés à un niveau global permettent de bien reproduire les quatre classes observées d'utilisation du titre. Par contre la variabilité inter-individuelle observée précédemment (figure 6.6 page 118) n'est pas bien simulée par notre modèle, car les différentes instances du modèle montrent peu de variabilité (voir figure 6.17 page ci-contre). Cette variabilité inter-individuelle est intéressante à simuler même si jusqu'à présent nous nous sommes attachés à reproduire les trajets d'un utilisateur moyen. Certains participants n'utilisent quasiment jamais le titre alors que d'autres lui font totalement confiance en passant de titre en titre.

Chaque instance du modèle a, au niveau intra-bloc, les mêmes paramètres. Ces paramètres sont les courbes décrites précédemment indiquant, suivant l'amplitude de la saccade, la probabilité de fixer ou non le titre. Pour simuler les différentes stratégies d'utilisation du titre observées chez les participants, une possibilité serait de faire varier ces courbes, comme illustré figure 6.18 page suivante. Ainsi pour

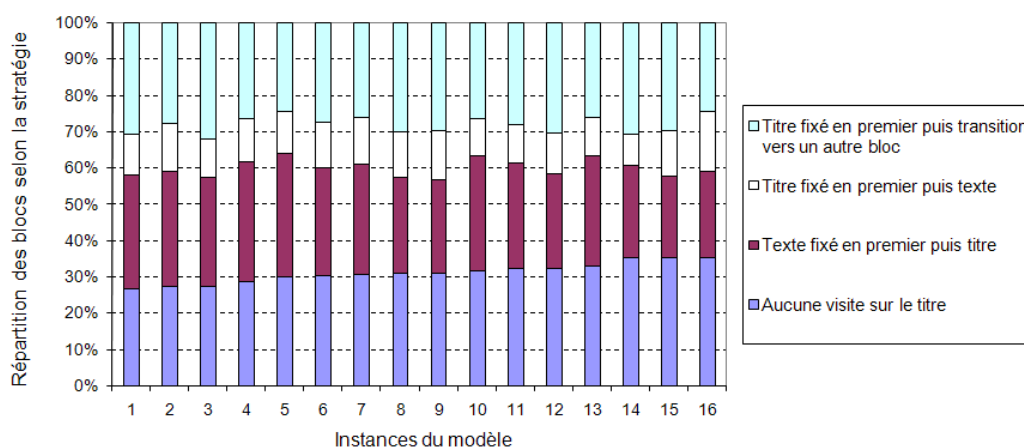


FIGURE 6.17 – Utilisation du titre dans les simulations.

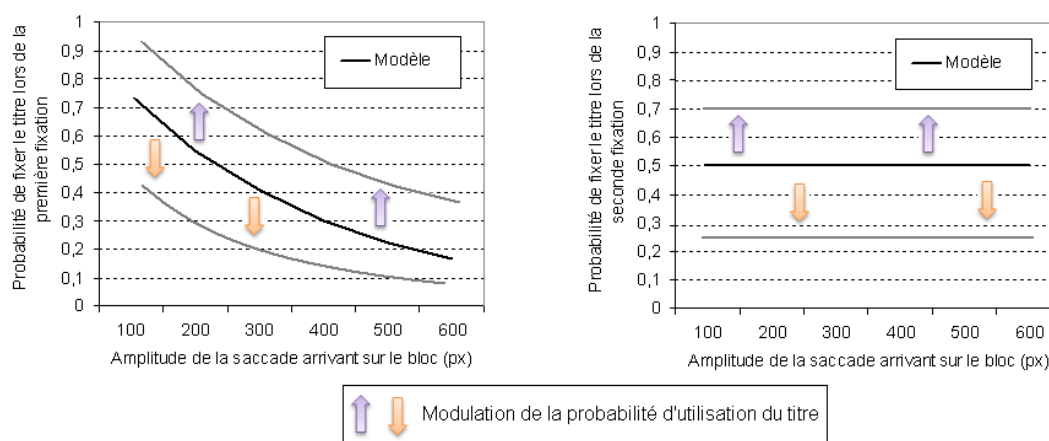


FIGURE 6.18 – Modulation possible de l'utilisation du titre dans les simulations.

simuler une utilisation très rare du titre, la courbe doit être baissée et pour simuler une utilisation quasiment systématique du titre cette courbe doit être relevée. La seconde probabilité, qui détermine si la seconde fixation sera sur le titre lorsque le texte a été fixé lors de la première fixation pourrait, elle aussi, varier selon les instances du modèle afin de reproduire les différences inter-individuelles observées.

6.4 Discussion

Dans ce chapitre nous avons vu quels étaient les processus nécessaires à la réalisation de la tâche, au niveau intra-bloc. À ce niveau, il s'agit d'un processus de lecture incomplète, avec une prise d'information particulière pour le titre. Cette lecture est incomplète car dans le cadre d'une recherche d'information, il est nécessaire de ne pas continuer la lecture lorsque celle-ci est en contradiction avec l'objectif recherché. La simulation que nous avons faite du processus de lecture est relativement simple par

rapport aux modèles de lecture existants (comme *EZ-Reader* ou SWIFT). En effet, même si nous avons tenu compte de la nature grammaticale des mots, notre modèle n'est pas capable, par exemple, de faire des saccades régressives, pourtant observées chez les lecteurs humains. Il n'est pas non plus capable à ce stade de sauter des mots, et ne tient pas compte de la longueur des mots. De plus les durées de fixations ne sont pas simulées.

La modélisation que nous proposons et qui a fait l'objet de deux publications (Chanceaux, Guérin-Dugué, Lemaire, & Baccino, 2009b, 2009c), permet de simuler correctement les performances moyennes des participants. Le traitement du titre, qui a un rôle privilégié tant visuel que sémantique reproduit le comportement d'un utilisateur moyen. Au niveau visuel, il y a une mise en évidence au-dessus du paragraphe, avec un mot isolé et donc saillant qui attire les fixations, et au niveau sémantique le titre porte une information pertinente par rapport au texte qu'il accompagne, et a donc un poids plus important dans les calculs de similarité. Le taux de lecture des paragraphes fortement liés au thème reste à améliorer, mais l'ajout de différentes stratégies de parcours intra-bloc (fixations *Exploratoires*), ou une modification des seuils de lecture dans le cas du titre, devrait réduire cette différence de comportement.

Dans l'analyse des données expérimentales nous avons repéré qu'un grand nombre de fixations étaient de type *Exploratoire*, alors que le modèle ne fait que de la lecture en séquence dans sa version actuelle. Une amélioration du modèle serait donc de simuler les deux types de fixations : *Lecture* et *Exploratoire*. Pour ce faire, l'idée serait de prendre en compte la longueur des mots, et leur saillance visuelle. Ainsi, si un mot dans le paragraphe est plus saillant que les autres il attirerait plus facilement les fixations. De plus le modèle par ces différents paramétrages rendra compte d'une propension plus ou moins importante de suivre telle ou telle stratégie, en lecture, ou en exploratoire, ou de manière plus réaliste en mélangeant des deux. Dans le cas où la stratégie exploratoire serait prédominante, le modèle irait de mot en mot dans le paragraphe mais sans suivre un schéma en séquence. Au niveau du traitement sémantique du paragraphe, le principe serait le même ; le modèle utiliserait la similarité (LSA) entre les mots traités (non linéairement cette fois) et le thème de la recherche pour décider s'il continue son exploration du bloc ou s'il le quitte. Nous pourrions aussi combiner les deux types de stratégie à l'intérieur d'un même bloc, avec une première partie *Exploratoire*, puis une partie *Lecture*, si par exemple le texte semble pertinent pour la recherche (même principe que pour le titre actuellement).

Par rapport à la variabilité inter-individuelle observée, ce continuum de profils dépend aussi certainement des différences d'expertise entre les participants et de l'homogénéité de l'échantillon, ainsi que des différences de prise en compte des moyens d'efficacité présents sur une page (organisation de la page, titre, mise en relief de mots, etc). De plus le nombre de participants dans la modalité *Avec Titre* (quinze) est peut-être insuffisant pour tirer des conclusions sur cette disparité. Une analyse plus approfondie de ces stratégies pourrait être utile afin de pouvoir les intégrer dans

les simulations.

Le type de traitement que nous avons modélisé au niveau du titre devrait pouvoir s'appliquer à d'autres éléments des pages web, comme les menus par exemple. Le modèle pourrait ainsi être appliqué à des pages plus riches et complexes.

Conclusion et perspectives

Discussion générale

Dans le cadre de cette thèse, nous avons ébauché une modélisation des processus impliqués dans la recherche d'information. La modélisation de ces processus s'est basée autant que faire se peut sur des théories et des hypothèses cognitives reconnues dans la communauté. Nous avons cherché dans cette étude à exprimer les processus cognitifs avec une exigence de simplicité et de généralité afin de les implémenter et d'augmenter leur pouvoir prédictif. Ces trois processus correspondent à des fonctions visuelles, sémantiques et mnésiques.

Le modèle prend en entrée une page contenant des informations visuelles et sémantiques ainsi qu'une consigne (un but de recherche, sous forme d'une question, ou d'un thème). Les sorties de notre modèle sont des trajets simulant les mouvements oculaires humains, c'est-à-dire des coordonnées XY sur la page.

Les différentes analyses que nous avons menées pour comparer les trajets issus du modèle et ceux produits par des participants nous ont permis d'une part d'améliorer le modèle et d'en fixer les paramètres et d'autre part de vérifier sa robustesse. Notre modèle a été testé sur différents stimuli, et a permis de simuler le comportement des utilisateurs. Les résultats que nous avons obtenus sont encourageants, et nous verrons à la fin de cette conclusion les perspectives de recherche envisageables.

Cette thèse s'est déroulée en collaboration avec Thierry Baccino et plusieurs étudiants de Master ont aussi participé à ce travail, notamment pour la passation des expériences.

Ma contribution a été la suivante :

Expérience avec les mots :

- Conception du matériel
- Passation de la version *Taille*
- Analyse des données
- Modélisation
- Comparaisons des données expérimentales et simulées

Expérience en lignes et en colonnes :

- Participation à la conception du matériel
- Analyse des données
- Modélisation
- Comparaisons des données expérimentales et simulées

Expérience en pêle-mêle :

- Participation à la conception du matériel
- Analyse des données
- Modélisation
- Comparaisons des données expérimentales et simulées

Nous allons revenir ici sur les étapes de modélisation ainsi que sur les différents paramètres utilisés.

Phases de paramétrisation du modèle

Dans une première étape, nous avons mis au point un modèle capable de simuler les trajets oculomoteurs sur une tâche simple de recherche d'information, sans traitement sémantique élaboré, dans des constellations de mots. Lors de cette première phase de travail, les paramètres suivants ont été utilisés :

- α_V , α_M et α_S : ce sont les poids relatifs des trois processus. Ils ont été fixés grâce aux données expérimentales.
- θ et Φ : ce sont les probabilités du processus mnésique. Elles ont été ajustées aux données expérimentales. Nous retrouvons les mêmes valeurs que celles de la littérature.
- δ : ce paramètre est utilisé dans l'approximation du processus de mémoire.
- seuil LSA : cette valeur est fixée à 0,2 d'après la littérature.

Nous avons ensuite complexifié la tâche et le matériel, afin de tester cette première version simple du modèle. La seconde version a été appliquée à deux organisations différentes de la page, les paragraphes étaient soit organisés en lignes et en colonnes soit en pêle-mêle (nous considérerons ici les paragraphes sans titre). Dans cette seconde version les paramètres sont les suivants :

- α_V et α_{MS} : ce sont les poids relatifs des trois processus (deux processus ont été regroupés). Les valeurs restent celles optimisées dans la version précédente.
- δ : ce paramètre est utilisé dans l'approximation du processus de mémoire. Il est identique à la version précédente.

- α_H et α_G : ces deux paramètres permettent de modéliser la stratégie globale de navigation dans la page. Ils reflètent les préférences humaines de parcours des documents (vers la droite et vers le bas). Ils ont été optimisés sur l'organisation en lignes et en colonnes et validés sur l'organisation pêle-mêle.
- seuils LSA : ces valeurs permettent de stopper ou non la lecture d'un paragraphe (seuil pour les paragraphes pertinents et non pertinents). Elles ont été estimées empiriquement et dépendent du nombre de mots traités.

Nous voyons dans cette seconde version du modèle que la plupart des paramètres viennent de l'expérience préalable. De plus les paramètres qui n'étaient pas dans la version précédente ont été optimisés sur l'organisation en lignes et en colonnes, ce qui nous a permis de réellement tester le modèle sur l'organisation en pêle-mêle où aucune optimisation n'a été effectuée.

Pour la modalité *Avec Titre* de la dernière expérimentation, nous avons rajouté des paramètres propres au traitement du titre :

- α_V et α_{MS} : ce sont les poids relatifs des trois processus (deux processus ont été regroupés). Les valeurs restent celles optimisées dans la version précédente.
- δ : ce paramètre est utilisé dans l'approximation du processus de mémoire. Il est identique à la version précédente.
- α_H et α_G : ces deux paramètres permettent de modéliser la stratégie globale de navigation dans la page. Ce sont les mêmes que dans la version précédente.
- seuils LSA : ces valeurs permettent de stopper ou non la lecture d'un paragraphe. Ce sont les mêmes que précédemment.
- seuil LSA pour le titre : cette valeur est fixée à 0,2 d'après la littérature.
- probabilités de fixer le titre : ces deux probabilités (pour la première et la seconde fixation) ont été fixées à partir des données empiriques en se basant néanmoins sur des hypothèses cognitives plausibles.
- poids du titre : ce paramètre donne deux fois plus de poids au titre dans le calcul de la similarité entre la portion de texte lue et le thème pour exprimer la prégnance de l'information du titre. C'est peut-être le moins solide de nos paramètres, qu'une nouvelle expérimentation permettrait de tester.

Les paramètres pour traiter le titre sont peu nombreux et la mise en œuvre des mécanismes de traitement du titre reste relativement simple. Suite à ce choix de simplicité, nous avons pu mettre en lumière certaines limites, que nous avons esquissées au fur et à mesure que nous les avons rencontrées et que nous allons résumer ici.

Tout d'abord le modèle n'est pas capable de stopper sa recherche. Cette fonctionnalité dépasse les ambitions actuelles du modèle, mais il serait néanmoins intéressant de prévoir un module de décision permettant l'arrêt en fin de recherche. Deuxièmement, notre matériel expérimental malgré une augmentation de sa complexité reste éloigné de la réalité. Nous avons fait ce choix afin de bien contrôler le matériel, et d'avancer pas à pas dans la construction du modèle. Chaque paramètre a donc été validé sur plusieurs expériences se complexifiant au fur et à mesure. Enfin, nous

avons voulu faire un modèle d'un utilisateur moyen, mais en observant plus finement les comportements des participants à nos expériences nous avons relevé une grande variabilité inter-individuelle, qui pourrait révéler des différences de stratégie, d'expertise, etc.

Perspectives de recherche

Nous avons relevé trois principales limites de notre modèle, les perspectives de recherche présentées ici s'appuient donc sur ces limites.

Une amélioration du modèle serait de lui permettre de s'arrêter à la fin de la recherche, et de ne pas lui fixer le nombre de blocs à visiter. Les taux de revisites que nous observons dans la version actuelle du modèle sont proches de ceux des participants, mais qu'en serait-il si le modèle avait un processus décisionnel lui permettant de prendre la décision d'arrêter la recherche ? Et est-ce que le nombre de blocs visités serait lui aussi comparable ? Les résultats que nous obtenons au niveau du parcours entre les blocs sont donc prometteurs, mais entraînent de nouvelles questions.

Au niveau de la complexité du matériel, nous ne pouvons que reconnaître que les pages Web dans lesquelles nous cherchons quotidiennement des informations sont beaucoup plus riches. Même si nous estimons que les processus que nous avons modélisés peuvent dans une certaine mesure s'appliquer à des environnements plus complexes, nous n'en avons pas de preuve tangible. De plus, il reste des éléments que nous n'avons pas du tout pris en compte comme les animations qui sont de plus en plus présentes sur nos écrans, la dynamique des pages Web (*scrolling*, etc), ainsi que la navigation entre les pages. Ces éléments se rajoutent aux éléments tels que les boutons, les menus, les liens, qui pourraient être intégrés au fur et à mesure dans le modèle. Il serait aussi intéressant de trouver une manière simple et automatique de traiter les photos.

Pour ce qui est de la variabilité inter-individuelle, une perspective de recherche serait de bien déterminer les différentes plages de variation des comportements et les stratégies sous-jacentes. Une piste de modélisation serait par exemple de simuler les deux types de fixations (*Lecture* et *Exploratoire*) et de pouvoir moduler leurs poids respectifs. Le modèle aurait ainsi la capacité de diversifier son comportement pour reproduire l'étendue des comportements observés et mieux comprendre ces différences inter-individuelles.

Enfin, suite au projet ANR Eye-LSA dans lequel s'inscrit cette thèse, un nouveau projet a pris le relais : le projet ANR Gaze&EEG auquel participent les laboratoires grenoblois Gipsa-Lab, TIMC et LPNC, et le laboratoire parisien LUTIN. Dans ce projet, un doctorant va continuer à travailler sur ces questions. Il disposera de données électro-encéphalographiques en plus des données oculométriques, ce qui lui permettra d'avoir accès à d'autres marqueurs des activités mentales pour enrichir le modèle.

Annexes

A Mots utilisés pour les expériences

Liste de toutes les définitions et de tous les mots triés selon leurs similarités sémantiques LSA avec la question posée.

Aliment le plus gras :		Aliment le plus sucré :		Animal le plus dangereux :		Animal le plus gros :		Animal le plus petit :	
mot-cible : beurre	mot-cible : confiture	mot-cible : tigre	mot-cible : baleine	mot-cible : tigre	mot-cible : baleine	mot-cible : crapaud	mot-cible : crapaud	mot-cible : tigre	mot-cible : crapaud
0.655885	0.561193	renard	0.595102	renard	0.553563	0.531819	0.531819	0.553563	0.531819
0.606078	0.488078	girafe	0.566599	girafe	0.552922	0.45448	0.45448	0.552922	0.45448
0.506525	0.450508	tigre	0.557777	tigre	0.418953	0.414645	0.414645	0.418953	0.414645
0.474421	0.429759	lièvre	0.534246	lièvre	0.498461	0.400067	0.400067	0.498461	0.400067
0.470234	0.419021	chien	0.496909	chien	0.430666	0.358761	0.358761	0.430666	0.358761
0.45732	0.389816	vache	0.454137	vache	0.404667	0.312748	0.312748	0.404667	0.312748
0.45259	0.38902	singe	0.441257	singe	0.389746	0.303455	0.303455	0.389746	0.303455
0.442177	0.354173	carapace	0.435495	carapace	0.388463	0.287825	0.287825	0.388463	0.287825
0.436738	0.343085	niche	0.413037	niche	0.356113	0.279695	0.279695	0.356113	0.279695
0.431304	0.32348	fermier	0.393998	fermier	0.349743	0.262267	0.262267	0.349743	0.262267
0.410439	0.312686	ventre	0.37912	ventre	0.334926	0.257899	0.257899	0.334926	0.257899
0.388631	0.296018	instinct	0.370475	instinct	0.321043	0.240681	0.240681	0.321043	0.240681
0.373614	0.285348	gueule	0.349602	gueule	0.292184	0.21697	0.21697	0.292184	0.21697
0.359291	0.271129	espèce	0.335113	espèce	0.288517	0.197932	0.197932	0.288517	0.197932
0.333565	0.261728	aliment	0.320729	aliment	0.284812	0.185991	0.185991	0.284812	0.185991
0.308811	0.246303	mètre	0.300579	mètre	0.279082	0.16667	0.16667	0.279082	0.16667
0.280563	0.23798	noisette	0.275532	noisette	0.266997	0.154946	0.154946	0.266997	0.154946
0.271413	0.225795	caverne	0.258695	caverne	0.244855	0.150195	0.150195	0.244855	0.150195
0.263268	0.210908	habitant	0.239905	habitant	0.229049	0.14588	0.14588	0.229049	0.14588
0.24705	0.1986	poison	0.232691	poison	0.215183	0.143415	0.143415	0.215183	0.143415
0.238746	0.193209	collier	0.227195	collier	0.209965	0.134941	0.134941	0.209965	0.134941
0.227376	0.189006	buisson	0.21724	buisson	0.196643	0.132691	0.132691	0.196643	0.132691
0.216513	0.183526	peste	0.207892	peste	0.187723	0.128076	0.128076	0.187723	0.128076
0.198058	0.172081	bosse	0.198009	bosse	0.181293	0.122397	0.122397	0.181293	0.122397
0.183781	0.155726	coquin	0.191479	coquin	0.174963	0.117594	0.117594	0.174963	0.117594
0.171011	0.14639	clown	0.1781	clown	0.163162	0.113163	0.113163	0.163162	0.113163
0.154517	0.127779	ennemi	0.166834	ennemi	0.16161	0.106624	0.106624	0.16161	0.106624
0.148411	0.114835	squelette	0.162584	squelette	0.148471	0.088373	0.088373	0.148471	0.088373
0.136731	0.104041	flèche	0.14927	flèche	0.130137	0.086085	0.086085	0.130137	0.086085
0.127141	0.093761	menace	0.133062	menace	0.126309	0.080786	0.080786	0.126309	0.080786
0.121266	0.083029	sursaut	0.120689	sursaut	0.117837	0.067641	0.067641	0.117837	0.067641
0.115293	0.073961	éclair	0.11101	éclair	0.108365	0.058107	0.058107	0.108365	0.058107
0.104453	0.060222	complice	0.100007	complice	0.094877	0.045342	0.045342	0.094877	0.045342
0.097793	0.048862	wagon	0.083175	wagon	0.073171	0.039524	0.039524	0.073171	0.039524
0.076473	0.051986	contraire	0.065061	contraire	0.063024	0.031074	0.031074	0.063024	0.031074
0.055385	0.047924	analogie	0.055649	analogie	0.057453	0.026907	0.026907	0.057453	0.026907
0.041669	0.038158	décal	0.046709	décal	0.046869	0.022734	0.022734	0.046869	0.022734
0.034439	0.021856	recul	0.019272	recul	0.036178	0.012323	0.012323	0.036178	0.012323
0.024667	0.018054	asile	0.006302	asile	0.028759	0.010233	0.010233	0.028759	0.010233
0.011333	0.007866	apprenti	0.001989	apprenti	0.005229	0.003919	0.003919	0.005229	0.003919

Arme la plus destructrice :		Couleur la plus sombre :		Instrument de musique le plus petit :		Métier le plus dangereux :		Meuble le mieux adapté au rangement :	
mot-cible : bombe	mot-cible : marron	mot-cible : flûte	mot-cible : soldat	mot-cible : armoire	mot-cible : marron	mot-cible : flûte	mot-cible : soldat	mot-cible : armoire	mot-cible : armoire
pistolet	0,44	jaune	0,831573	violon	0,562326	ouvrier	0,437167	chaise	0,422773
revolver	0,418995	rouge	0,795893	guitare	0,553381	maton	0,305363	canapé	0,392848
carabine	0,39358	violet	0,721093	piano	0,54907	artisan	0,27555	fauteuil	0,375579
fusil	0,368626	orange	0,6616	trompette	0,42049	soldat	0,259542	armoire	0,367583
bombe	0,31192	blanc	0,610449	batterie	0,40059	artiste	0,258486	table	0,346116
sabre	0,294812	marron	0,519574	flûte	0,382273	sculpteur	0,245465	bureau	0,205335
couteau	0,261841	pourpre	0,516043	tambour	0,371168	marchand	0,2419	coffre	0,200282
accident	0,259788	feuillage	0,478799	orchestre	0,362337	citoyen	0,222879	assiette	0,192959
cadavre	0,256049	peinture	0,472935	rythme	0,344865	anecdote	0,195394	linge	0,191024
blessure	0,241749	tableau	0,430218	timbre	0,333622	chantier	0,190634	rideau	0,191174
fuite	0,230069	relief	0,425297	chanteuse	0,329137	maladie	0,183447	colonne	0,18816
assassin	0,224288	portrait	0,417702	thème	0,309521	vapeur	0,181236	gâteau	0,185872
torture	0,212771	pantalon	0,409536	disque	0,296625	loisir	0,174316	comtesse	0,18357
veste	0,203638	fleur	0,383707	tradition	0,28041	manuel	0,173706	invention	0,178726
moteur	0,190504	dessin	0,375196	luthin	0,276785	marine	0,169379	ménage	0,176275
destin	0,180148	visage	0,368632	rossignol	0,27135	cuisinier	0,164869	cuisinier	0,174169
cible	0,171389	bouquet	0,355748	album	0,251692	ruine	0,162634	beauté	0,158894
punition	0,155157	lumière	0,353042	oreille	0,24794	plaine	0,160511	mousse	0,152581
froid	0,146672	encre	0,344127	singe	0,23596	mareau	0,150047	verdure	0,145158
vainqueur	0,123462	parfum	0,335491	crapaud	0,214963	engin	0,137862	signe	0,141276
hôpital	0,119049	amateur	0,310146	rose	0,206229	oeuvre	0,135906	cravate	0,137163
légende	0,106388	habit	0,294495	prairie	0,203414	dictature	0,133453	volet	0,133029
filet	0,098622	étoffe	0,276713	bague	0,195164	institut	0,131453	bouquet	0,126657
photo	0,092365	papier	0,270779	panier	0,190071	vigne	0,125082	fromage	0,114753
épisode	0,086106	vigne	0,250591	opéra	0,18705	imitation	0,121163	commerce	0,110408
chasseur	0,082435	foulard	0,226731	audition	0,176903	misère	0,119179	angle	0,098948
confort	0,080333	spectre	0,194644	registre	0,164387	accident	0,112433	brosse	0,094892
balance	0,071901	galerie	0,184355	nostalgie	0,152315	geste	0,102075	cabinet	0,092833
teint	0,063554	noeud	0,157015	village	0,143951	exception	0,090833	client	0,090694
argument	0,046924	apprenti	0,151394	tumulte	0,133373	tutelle	0,081	partition	0,080896
cantine	0,042746	branche	0,138664	notable	0,119607	agitation	0,073534	vache	0,066811
vacarme	0,041246	beauté	0,126751	pureté	0,112236	impôt	0,056321	besoin	0,064268
cabane	0,034935	soleil	0,093552	dragon	0,10185	cravate	0,05046	veste	0,049401
emblème	0,032853	dimension	0,091186	culture	0,088415	appareil	0,040639	basket	0,037481
banque	0,030729	complot	0,086647	vacarme	0,075781	montage	0,032888	millier	0,034401
anecdote	0,02644	légion	0,070478	copie	0,066174	visage	0,030917	retrait	0,032371
hausse	0,024359	couple	0,051131	richesse	0,047349	signe	0,026964	arsenal	0,030343
versant	0,018441	quartier	0,031854	addition	0,035038	accueil	0,024997	dessein	0,026166
terrier	0,007911	chanteuse	0,015043	insecte	0,017024	blaireau	0,013495	choeur	0,016186
version	0,005845	attente	0,006367	gâteau	0,011867	linge	0,009395	tyran	0,001215

Plante la plus piquante :		Poisson le plus dangereux :		Récipient ayant la plus grande contenance :		Sport le plus brutal :		Boisson la plus alcoolisée :	
mot-cible : cactus	mot-cible : requin	mot-cible : marmite	mot-cible : rugby	mot-cible : liqueur	mot-cible : requin	mot-cible : marmite	mot-cible : rugby	mot-cible : liqueur	mot-cible : liqueur
0,709028	0,731732	0,749805	0,711557	0,827427	baleine	casserole	football	tisane	0,827427
0,628577	0,666712	0,585464	0,685904	0,74313	anguille	poêle	rugby	sirop	0,74313
0,589032	0,646103	0,544115	0,573099	0,686888	requin	assiette	tennis	liqueur	0,686888
0,579629	0,595384	0,482028	0,481878	0,670755	truite	verre	basket	vinaigre	0,670755
0,571925	0,542413	0,447422	0,429139	0,646507	saumon	bouteille	cyclisme	limonade	0,646507
0,546718	0,534672	0,416811	0,422025	0,310491	dauphin	marmite	course	bière	0,310491
0,526526	0,451893	0,359173	0,378284	0,293403	brochet	tasse	voile	cidre	0,293403
0,518267	0,438996	0,345962	0,345961	0,287083	mouette	pomme	gardien	raisin	0,287083
0,479042	0,435962	0,340885	0,360459	0,282212	navire	boullanger	victoire	carafe	0,282212
0,470848	0,418883	0,335915	0,352605	0,250255	rivage	agneau	vainqueur	saveur	0,250255
0,464259	0,402764	0,335705	0,320304	0,240562	barque	couteau	journal	chardon	0,240562
0,457366	0,391457	0,333269	0,313777	0,238005	sable	alcool	fracture	poison	0,238005
0,418429	0,362524	0,331414	0,306235	0,235516	queue	gâteau	tribune	tabac	0,235516
0,388349	0,355339	0,327895	0,294396	0,224871	cargaison	boisson	record	boullanger	0,224871
0,385191	0,345748	0,313465	0,29239	0,216305	corail	poudre	épreuve	peste	0,216305
0,366211	0,338571	0,29688	0,27637	0,211082	faune	cuisine	poule	rougeur	0,211082
0,354778	0,323198	0,254451	0,258149	0,202975	bouillon	savon	canitine	privation	0,202975
0,344903	0,305368	0,238535	0,241767	0,196052	plumage	fumée	amateur	sommeil	0,196052
0,327569	0,296236	0,232546	0,211736	0,185219	beurre	poisson	effort	bouchon	0,185219
0,308821	0,286451	0,210678	0,20079	0,18246	radeau	flamme	boulot	piqûre	0,18246
0,300214	0,276791	0,188266	0,191225	0,177311	brochette	odeur	drapeau	tuyau	0,177311
0,277968	0,25629	0,178483	0,185521	0,166722	rivière	carré	antenne	bassin	0,166722
0,254848	0,253876	0,15673	0,180279	0,156298	fleuve	fonte	sifflet	danger	0,156298
0,241425	0,23436	0,152015	0,17507	0,144959	saucisse	dragon	leçon	prunier	0,144959
0,225676	0,226194	0,136823	0,156649	0,139985	dindon	marteau	budget	filet	0,139985
0,206747	0,189005	0,132535	0,136189	0,121356	morceau	poignée	cahier	pommier	0,121356
0,19928	0,17543	0,116078	0,129512	0,112173	framboise	abondance	titre	sourde	0,112173
0,189073	0,168192	0,099497	0,122841	0,10684	saive	vitre	manège	cocaine	0,10684
0,181443	0,158103	0,090816	0,118208	0,100048	renard	marchande	bateau	moustique	0,100048
0,172057	0,14822	0,084231	0,103645	0,08729	colonie	origine	chanteuse	carte	0,08729
0,157203	0,144148	0,074542	0,092636	0,080455	martien	chevre	réserve	ombrage	0,080455
0,144547	0,128502	0,061967	0,084982	0,071499	chasse	artifice	coque	tricot	0,071499
0,125282	0,120783	0,061043	0,069675	0,067111	divorce	muscle	vingtaine	fourche	0,067111
0,110055	0,105679	0,050455	0,059041	0,058865	calcaire	maximum	légende	agonie	0,058865
0,096346	0,087548	0,044247	0,048233	0,050047	cornet	éloge	oxygène	anomalie	0,050047
0,083801	0,077292	0,0422	0,036155	0,047856	galère	convoi	volonté	fascisme	0,047856
0,066286	0,073001	0,038465	0,019075	0,045672	copain	période	crendre	alinéa	0,045672
0,050647	0,045843	0,034298	0,016936	0,02591	festin	énigme	charbon	mouette	0,02591
0,030773	0,035592	0,018215	0,01474	0,02372	calomnie	arbre	chêne	relation	0,02372
0,013229	0,022954	0,001733	0,006178	0,010861	roulette	finesse	domicile	tarif	0,010861

Aliment le plus liquide :		Habit le mieux adapté pour aller dehors :		Instrument de musique le plus encombrant :		Transport le plus rapide :		Jouet le plus approprié pour un bébé :	
mot-cible : bière		mot-cible : manteau		mot-cible : orgue		mot-cible : avion		mot-cible : peluche	
sucre	0,535988	chemise	0,355113	guitare	0,725347	avion		peluche	
bière	0,441928	manteau	0,341943	violon	0,70194	train		poupée	
viande	0,441606	peignoir	0,292855	harpe	0,70116	voiture		tracteur	
farine	0,406553	pyjama	0,272843	orgue	0,6104011	autobus		nourrice	
beurre	0,352067	cravate	0,264633	trompette	0,574178	navire		canard	
poulet	0,345539	gilet	0,220905	flûte	0,524716	bateau		neveu	
salade	0,329135	culotte	0,216259	tambour	0,46678	camion		grelot	
consève	0,314952	glace	0,214435	timbre	0,447649	ligne		dauphin	
recette	0,302131	visage	0,200236	rythme	0,441088	panne		noisette	
adulte	0,270817	frisson	0,195255	thème	0,421541	course		gôûter	
noyau	0,268519	laine	0,19111	gamme	0,404566	voyageur		écharpe	
assiette	0,260708	jardin	0,172982	choeur	0,37234	budget		mariée	
tabac	0,246079	flambée	0,160948	variation	0,359968	réseau		hôpital	
plastique	0,237575	mouchoir	0,154948	refrain	0,315609	nauffrage		voiture	
blessure	0,22379	tempête	0,147235	bague	0,299271	météo		chaise	
chair	0,216807	sifflet	0,135998	échelle	0,263683	effort		fauteuil	
plaque	0,210207	brise	0,133992	audition	0,245019	béton		barreau	
hygiène	0,204247	épaule	0,121945	opéra	0,242362	courant		serveur	
matière	0,196879	contraire	0,11634	registre	0,235872	logiciel		éveil	
cuisine	0,194013	gôûter	0,092569	ambiance	0,216994	marteau		bonnet	
remède	0,178218	noeud	0,086484	notable	0,202682	voyage		foulard	
bouche	0,172741	boullanger	0,084328	poème	0,185163	brume		chocolat	
paille	0,152738	cabinet	0,079089	bruit	0,177981	brèche		pente	
trouble	0,140125	sommet	0,074942	effectif	0,175315	énergie		pantin	
médecin	0,13797	police	0,072898	amateur	0,166854	sentier		vache	
attaque	0,114833	pureté	0,070918	facture	0,162058	monopole		collier	
insecte	0,108238	buisson	0,068862	contexe	0,140551	rivière		cratère	
lumière	0,099487	merveille	0,055372	offrande	0,13689	boulevard		vacarme	
magistrat	0,094862	pistolet	0,053384	vigueur	0,134669	extension		saumon	
courant	0,083478	écume	0,051376	fossé	0,132307	santé		balade	
tache	0,079272	asile	0,043189	cassette	0,131551	cadre		bêtise	
carte	0,074963	politesse	0,039639	support	0,128876	fantôme		romah	
colonne	0,070681	paysage	0,037643	habitant	0,126216	régime		échelon	
époque	0,068279	hexagone	0,035629	volonté	0,118872	calcul		casserole	
pensée	0,0503	emblème	0,031543	canon	0,101748	troupeau		comédie	
explosion	0,041862	bouclier	0,025323	hygiène	0,032623	dessin		festival	
aimant	0,038432	criminel	0,021268	douleur	0,023508	poupée		marmite	
torture	0,032122	mélodie	0,01776	pluie	0,019826	remarque		portion	
éveil	0,016038	tronc	0,013771	violence	0,013019	sportif		variation	
satellite	0,007657	guide	0,005447	poignard	0,004103	accent		doute	

Aliment le plus liquide :		Habit le mieux adapté pour aller dehors :		Instrument de musique le plus encombrant :		Transport le plus rapide :		Jouet le plus approprié pour un bébé :	
mot-cible : bière		mot-cible : manteau		mot-cible : orgue		mot-cible : avion		mot-cible : peluche	
sucre	0,535988	chemise	0,355113	guitare	0,725347	avion		peluche	
bière	0,441928	manteau	0,341943	violon	0,70194	train		poupée	
viande	0,441606	peignoir	0,292855	harpe	0,70116	voiture		tracteur	
farine	0,406553	pyjama	0,272843	orgue	0,6104011	autobus		nourrice	
beurre	0,352067	cravate	0,264633	trompette	0,574178	navire		canard	
poulet	0,345539	gilet	0,220905	flûte	0,524716	bateau		neveu	
salade	0,329135	culotte	0,216259	tambour	0,46678	camion		grelot	
consève	0,314952	glace	0,214435	timbre	0,447649	ligne		dauphin	
recette	0,302131	visage	0,200236	rythme	0,441088	panne		noisette	
adulte	0,270817	frisson	0,195255	thème	0,421541	course		gôûter	
noyau	0,268519	laine	0,19111	gamme	0,404566	voyageur		écharpe	
assiette	0,260708	jardin	0,172982	choeur	0,37234	budget		mariée	
tabac	0,246079	flambée	0,160948	variation	0,359968	réseau		hôpital	
plastique	0,237575	mouchoir	0,154948	refrain	0,315609	nauffrage		voiture	
blessure	0,22379	tempête	0,147235	bague	0,299271	météo		chaise	
chair	0,216807	sifflet	0,135998	échelle	0,263683	effort		fauteuil	
plaque	0,210207	brise	0,133992	audition	0,245019	béton		barreau	
hygiène	0,204247	épaule	0,121945	opéra	0,242362	courant		serveur	
matière	0,196879	contraire	0,11634	registre	0,235872	logiciel		éveil	
cuisine	0,194013	gôûter	0,092569	ambiance	0,216994	marteau		bonnet	
remède	0,178218	noeud	0,086484	notable	0,202682	voyage		foulard	
bouche	0,172741	boullanger	0,084328	poème	0,185163	brume		chocolat	
paille	0,152738	cabinet	0,079089	bruit	0,177981	brèche		pente	
trouble	0,140125	sommet	0,074942	effectif	0,175315	énergie		pantin	
médecin	0,13797	police	0,072898	amateur	0,166854	sentier		vache	
attaque	0,114833	pureté	0,070918	facture	0,162058	monopole		collier	
insecte	0,108238	buisson	0,068862	contexe	0,140551	rivière		cratère	
lumière	0,099487	merveille	0,055372	offrande	0,13689	boulevard		vacarme	
magistrat	0,094862	pistolet	0,053384	vigueur	0,134669	extension		saumon	
courant	0,083478	écume	0,051376	fossé	0,132307	santé		balade	
tache	0,079272	asile	0,043189	cassette	0,131551	cadre		bêtise	
carte	0,074963	politesse	0,039639	support	0,128876	fantôme		romah	
colonne	0,070681	paysage	0,037643	habitant	0,126216	régime		échelon	
époque	0,068279	hexagone	0,035629	volonté	0,118872	calcul		casserole	
pensée	0,0503	emblème	0,031543	canon	0,101748	troupeau		comédie	
explosion	0,041862	bouclier	0,025323	hygiène	0,032623	dessin		festival	
aimant	0,038432	criminel	0,021268	douleur	0,023508	poupée		marmite	
torture	0,032122	mélodie	0,01776	pluie	0,019826	remarque		portion	
éveil	0,016038	tronc	0,013771	violence	0,013019	sportif		variation	
satellite	0,007657	guide	0,005447	poignard	0,004103	accent		doute	

B Questionnaire

Questionnaire qu'ont rempli 26 participants afin de faire un classement plus fin que celui de LSA sur les sept mots de la catégorie.

C Matrice de similarités

Matrice de similarités entre tous les mots d'une page. La question est *Quel est le poisson le plus dangereux ?*

D Pages pour les différentes questions de l'expérience avec les mots

Pages pour toutes les questions dans la modalité sémantique sans visuel (tous les mots de taille 15).

cantine
 photo torture
 punition
 assassin
 anecdote
 version
 hôpital
 accident cible épisode argument
 blessure revolver
 confort
 veste sabre bombe vainqueur moteur
 pistolet
 destin
 couteau fusil hausse
 carabine cadavre légende emblème
 fuite vacarme
 banque
 froid balance
 cabane chasseur versant
 teint terrier file

mètre recul
 habitant
 wagon
 niche asile délai
 girafe carapace
 buisson caverne flèche
 bosse collier éclair sursaut
 lièvre ventre complice menace
 gueule
 renard tigre fermier
 singe vache ennemi contraire
 chien noisette coquin poison instinct apprenti
 clown peste
 espèce
 squelette aliment analogie

Arme la plus destructrice

Animal le plus dangereux

sportif aliment squelette pêche
 variation
 graisse
 bétail laine assiette
 provision coquille
 instinct planche chocolat
 mutation
 singe baleine
 symbole canard cochon
 fondateur flanc ferme renard chien étable
 piège berger fermier tigre collier
 jungle buisson vaisselle
 poursuite cachette
 caresse
 appel
 besoin
 docteur

magistrat
 médecin
 remède blessure torture
 trouble
 adulte
 carte bouche attaque insecte
 salade poulet cuisine tabac tache éveil
 viande chair
 farine sucre bière conserve plaque aimant satellite
 beurre recette paille courant explosion
 assiette plastique hygiène colonne matière époque
 noyau
 pensée

Animal le plus gros

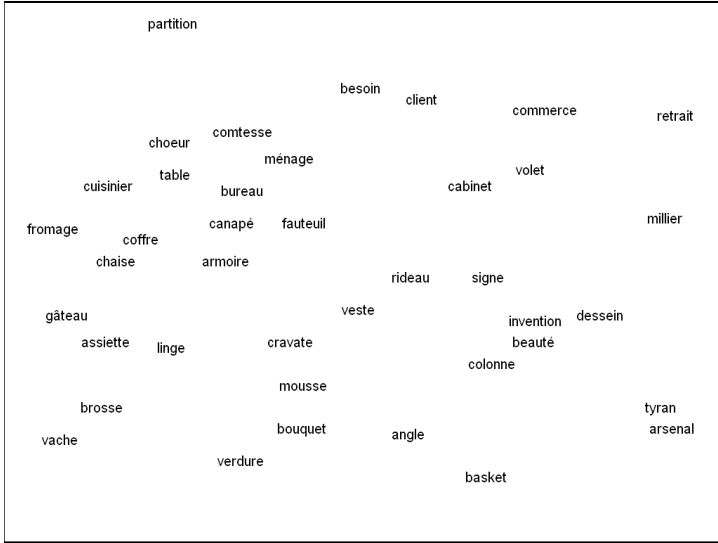
Aliment le plus liquide

exercice
 magistrat alarme élevage niche museau pelage
 ferme fourrure
 fontaine troupeau herbe
 carré grain tigre
 angle planche buisson chasseur renard
 crapaud
 empereur surprise chameau canard crocodile
 gaillard gamin
 noisette
 front fermier
 coude chaise casque meunier
 frère chocolat
 argent

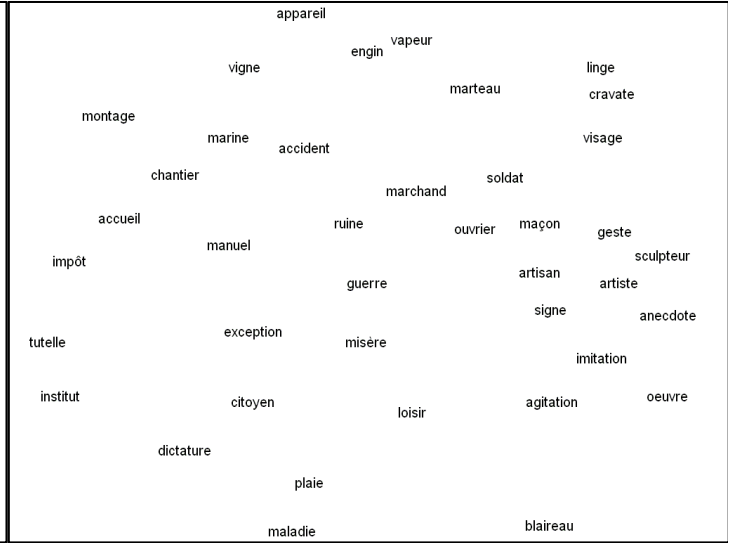
seigneur tyrannie
 pasteur oxygène
 parrain peste hygiène énergie
 viande fromage boulanger matière
 beurre sucre poudre sueur usage portion
 cuisson soupe poulet morceau défaut fertilité
 assiette
 saladade poêle paille couture
 botte
 fermier cuisine linge
 hôteesse
 blonde pièce olivier
 match
 amende football

Animal le plus petit

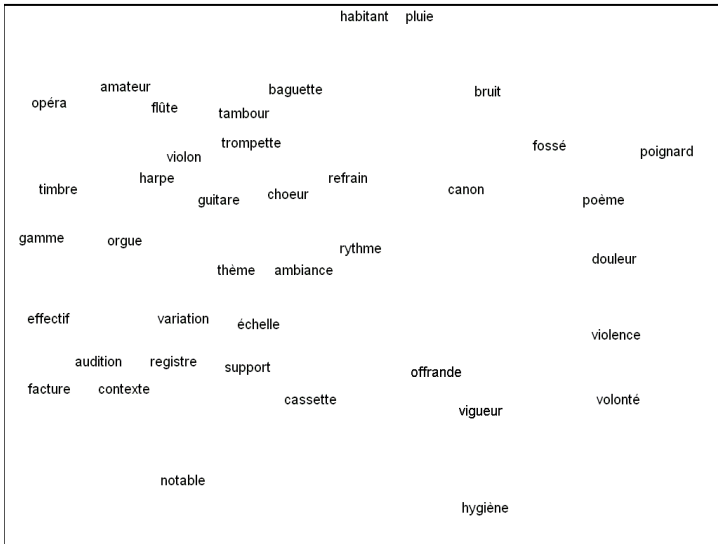
Aliment le plus gras



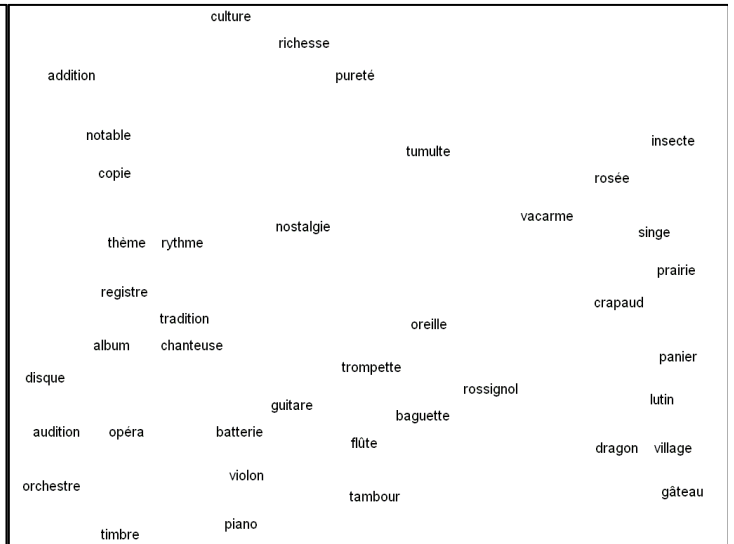
Meuble le mieux adapté au rangement



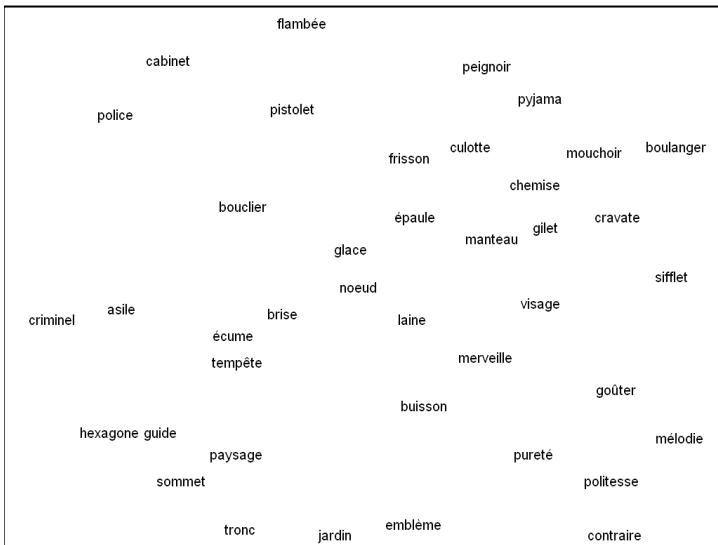
Métier le plus dangereux



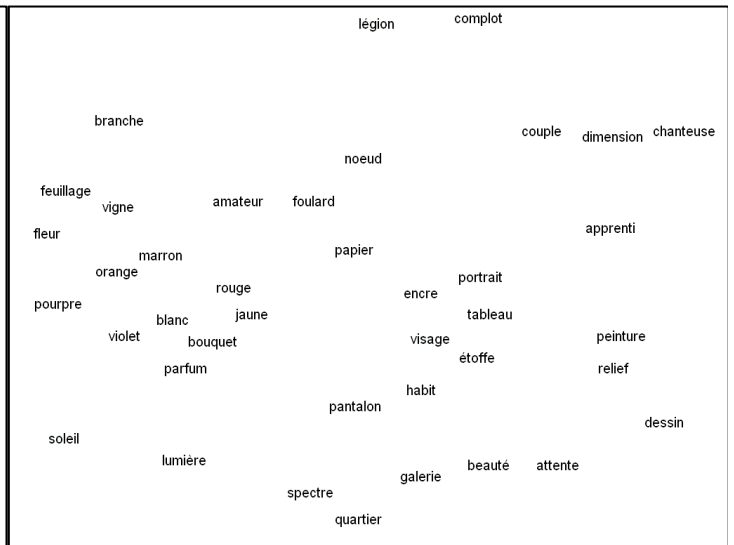
Instrument de musique le plus encombrant



Instrument de musique le plus petit



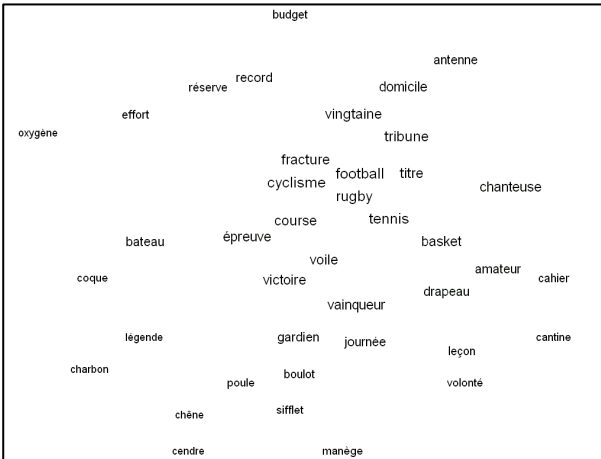
Habit le mieux adapté pour aller dehors



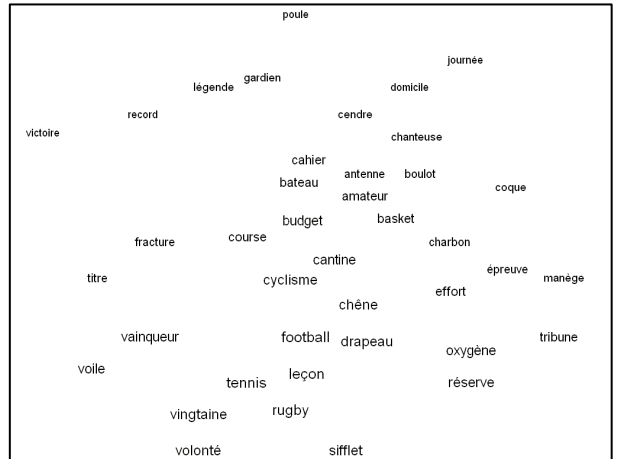
Couleur la plus sombre

E Pages pour les différentes modalités de l'expérience avec les mots

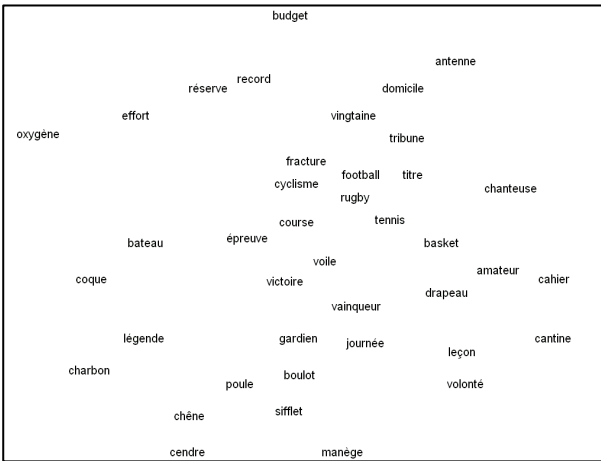
Pages obtenues pour la question *Quel est le sport le plus brutal ?* en croisant les modalités visuelles et sémantiques pour l'expérience dans la version *taille* et *couleur*.



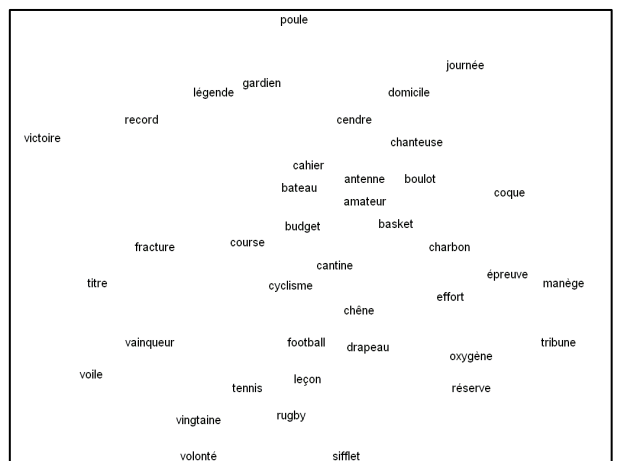
Sem V+



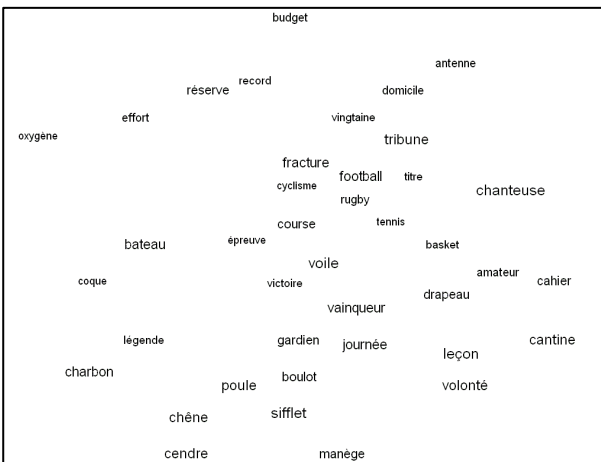
NoSem V+



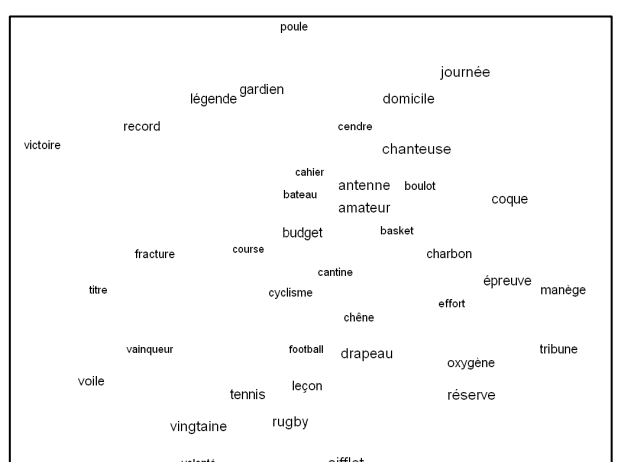
Sem V=



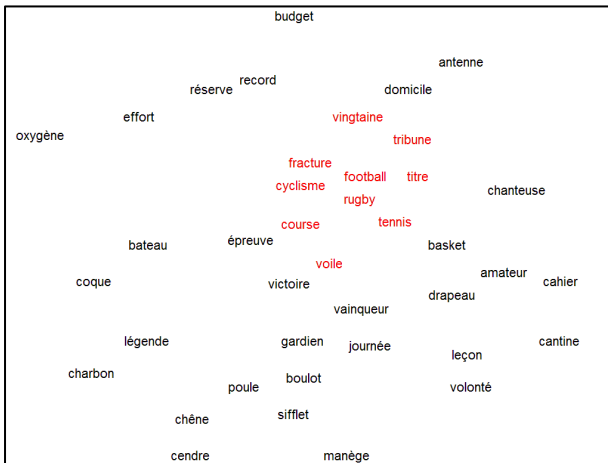
NoSem V=



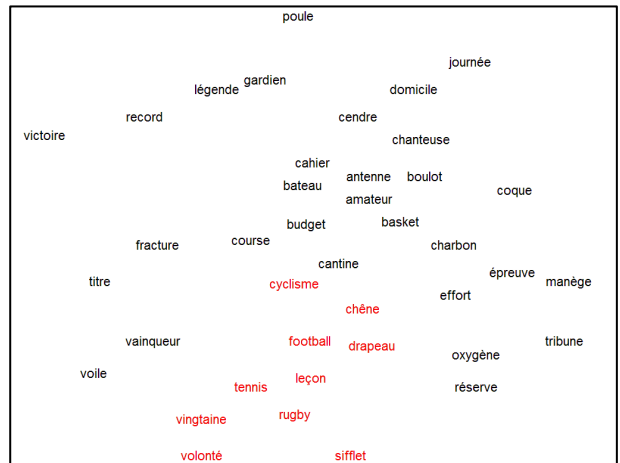
Sem V-



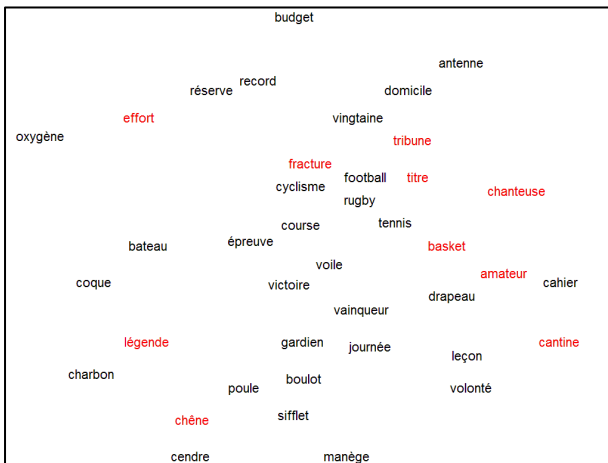
NoSem V-



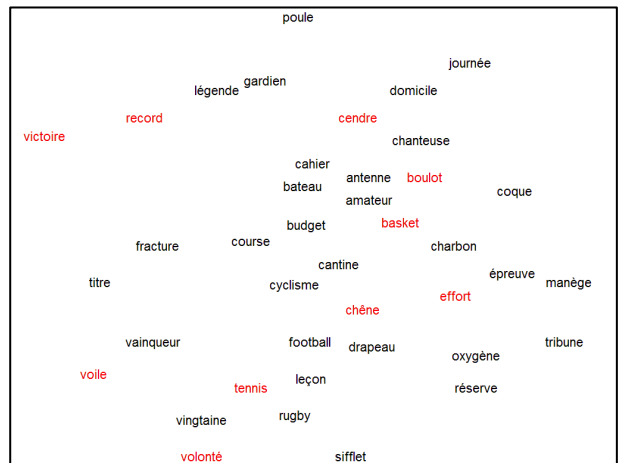
Sem V+



NoSem V+



Sem V-



NoSem V-

F Textes pour l'expérience avec les paragraphes

Textes pour tous les thèmes de l'expérience avec les paragraphes. Pour chaque thème les 2 premiers textes sont ceux avec une forte association sémantique (AF), les deux suivants ceux avec une association moyenne (Afa) et les trois derniers ceux sans aucune association (AA).

Thème	Type	Paragraphe
Aide aux réfugiés	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Associations humanitaires	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Chef de la Russie	AF	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	AF	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	Afa	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	Afa	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	AA	Il n'est point de bon mercato sans un feuilleton pour l'animer. Les transferts des deux footballeurs n'ont pas électrisés les foules en dehors du premier cercle des admirateurs de l'un et de l'autre.
Chef de la Russie	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	Le monocoque Garmore 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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Faiblesse du Dollar	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	La plupart des animateurs-moderateurs 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	AF	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 mises en place.
Formation en informatique	Afa	Bien plus que la connaissance d'une langue morte, l'apprentissage du latin dès le collège est un exercice de raisonnement et d'affirmation du caractère. Il est un puissant signe de différence et de condition supérieure.
Formation en informatique	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	Les transports connaissent 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	Afa	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	Afa	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	AA	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 maîtrisée.
Grève des cheminots	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Prix du pétrole	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Réchauffement climatique	AF	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	AF	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	Afa	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	Afa	Un groupe d'ingénieurs météorologistes a obtenu le prix Nobel pour ses travaux relatifs à la prévision des événements climatiques à long terme et aux effets de la pollution de l'atmosphère.

Réchauffement climatique	AA	Dans le jugement, cité par La Nación, les juges mettent en avant plusieurs arguments de forme, arguant que le type de procédure lancé par le juge Cerda serait inapproprié.
Réchauffement climatique	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	Le chanteur Yannick Noah, ici en concert, a repris sa place à la tête du classement des personnalités préférées des Français et devance Zinédine Zidane.
Réforme de la justice	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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.
Tourisme en montagne	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	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	Afa	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	Afa	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	AA	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	AA	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	AA	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	AF	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	AF	L'équipe de France de football a vaincu l'Australie en finale de la Coupe du Monde. L'entraîneur est très satisfait de son équipe et envisage les rencontres futures avec enthousiasme et sérénité.
Victoire des footballeurs	Afa	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	Afa	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	AA	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	AA	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	AA	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.

G Pages d'un thème pour l'expérience structurée en lignes et colonnes

Pages obtenues pour le thème *Croissance de l'économie* en alternant la place des blocs, celle de la photo et en changeant de blocs colorés.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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 lanterne passée sur le front.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 lanterne passée sur le front.

Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.

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 lanterne passée sur le front.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 lanterne passée sur le front.

Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 lanterne passée sur le front.

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.


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.

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.

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.

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.

Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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 lanterne passée sur le front.


Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.

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.

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.

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.

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.



H Pages d'une liste pour l'expérience structurée en lignes et colonnes

Pages obtenues pour tous les thèmes dans une liste.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

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.

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.



Les Etats-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.

L'arsenal militaire du pays s'est enrichi d'armes à micro-ondes susceptibles de détruire les circuits informatiques et les équipements de télécommunications. Il a ses militaires veulent les garder pour s'en servir contre les infrastructures ennemies.

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.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

Les élections de l'Iowa ont été sans doute les plus disputées de l'histoire de cet Etat rural. La participation à ces assemblées a dépassé tous les espoirs prévus par les politiques.

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.

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.

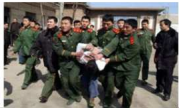


La déflagration s'est produite au passage d'un véhicule militaire sur une route du centre-ville, à une centaine de mètres d'une base de l'armée turque, selon la police.

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.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

Le monocoque Garmore du Britannique a décollé, dans la troisième étape de l'Around Alone, tour du monde à la voile en solitaire avec escales, entre Auckland et Punta del Este.

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.

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.

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.

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.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

A Moscou, les communistes triomphants se succèdent. Dans le Daghestan, les rebelles « sont complètement encerclés », leur chef militaire « est blessé », les islamistes ont subi de lourdes pertes.

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 enduits de curare.



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.

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.

Depuis le limogeage surprise du précédent chef du gouvernement russe pendant cinq ans de « stabilité », le président tente sans succès de maîtriser sa succession. Les prétendants sont nombreux.

Il n'est point de bon mercato sans un feuillet. 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.

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.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 d'avoir lancé des opérations armées.

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.

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.

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.

L'état de l'unique route qui mène 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.

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.



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.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

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.

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.

Le chef du gouvernement est accusé des relations tendues avec les Etats-Unis, flicage d'Israël dans le monde, le chômage en montée vertigineuse, la détérioration de toutes les normes de gouvernement démocratique.

L'artillerie israélienne a massivement bombardé des régions soupçonnées être des bastions du Hezbollah au Liban. Ces tirs sont une riposte à une série d'attaques meurtrières de la milice intégriste Chite contre Israël.

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 ses brevets dans le domaine public.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Un domaine de réflexion a été identifié. La référence est de plus en plus grande chez certains propriétaires de contenus à commercialiser en ligne par crainte du piratage et du téléchargement illégal.

La plupart des animateurs-moderateurs ont une formation de journaliste. Il n'existe pas de formation spécifique, mais les écoles de journalisme ou de communication qui donnent des cours en multimédia.


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.

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.

Chez les démocrates, les électeurs débattent de leurs choix respectifs et tendent de persuader leurs voisins de se rallier à leur candidat. Côté républicain, le scrutin se déroule rapidement après l'ouverture des assemblées.

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.

Bien plus que la connaissance d'une langue morte, l'apprentissage du latin dès le collège est un exercice de raisonnement et d'affirmation du caractère. Il est un puissant signe de différence et de condition supérieure.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Si la Californie avait appliqué sa politique de réduction des émissions de gaz, près de la moitié des véhicules vendus aux Etats-Unis auraient pu être concernés à moyen terme.

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.


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.

Avec l'aide de leurs rabatteurs déguillés, 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.

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.

Le skippeur rend hommage aux concepteurs. Le pan de la légèreté et de la simplicité était osé. Les gens qui se sont investis pour que ce projet existe peuvent être fiers.

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 %.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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.

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.

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.

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.

Cette réaction conduit industriels et milieux financiers à la prudence. D'autant que le champion des OGM a vu son chiffre d'affaires baisser. Les valeurs de biotechnologie ont perdu de leur attrait.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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 maîtrisée.

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.

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.

Les transports connaissent 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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Dans les clarières, 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.


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.

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.

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.

Les marges de croissance pourraient être augmentées par des réformes structurelles sur les marchés du travail et des biens. Accélération de la croissance, stabilité des prix, excédent de la balance des paiements courants.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

L'étude du système solaire est difficile. 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.

Laura 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.

Le retour à la vraie vie peut s'avérer douloureux pour ceux qui n'ont pas économisé les dollars pour gagner les matelas qui atténuent leur atterrissage. C'est le cas de ce basketteur célèbre.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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.

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.

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.

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.

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.



Plus sérieusement, cette saison, Tony Parker n'a plus confiance sur le terrain. Il y a des périodes pendant les matches où il comprend qu'il peut totalement dominer.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA


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.

Dans le jugement, cité par La Nación, les juges mettent en avant plusieurs arguments de forme, arguant que le type de procédure lancé par le juge Cerda serait inapproprié.

Un groupe d'ingénieurs météorologistes a obtenu le prix Nobel pour ses travaux relatifs à la prévision des événements climatiques à long terme et aux effets de la pollution de l'atmosphère.

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 produits par les rejets massifs de carbone associés aux activités touristiques.

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.



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.


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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 localisés aidés.

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.



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.

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.

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 ornées de charmants balcons.

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.


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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.



Le chanteur Yannick Noah ici en concert a repris sa place à la tête du classement des personnalités préférées des Français et devance Zinedine Zidane.

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.

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 entraîné des condamnations.

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.

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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA


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.

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.

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.

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.

La Californie a officiellement déposé plainte contre le gouvernement des Etats-Unis pour avoir empêché l'application de sa loi de réduction des émissions de gaz à effet de serre des automobiles.



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.

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.

Le Journal
Mise à jour à 07 h 03 - Paris


Regardez mieux pour mieux comprendre
EYELSA

Il faut profiter de la crise générale que traverse le gouvernement et de la fatigue de certaines unités militaires. Le guénilero juge nécessaire de mener des actions armées sur les routes.

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.

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.



Des trafiquants de cocaïne colombiens ont sabordé 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.



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.


Une semaine à Santorini, 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.


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.

Le Journal	Regardez mieux pour mieux comprendre EYELSA	Le Journal	Regardez mieux pour mieux comprendre EYELSA
<p>L'équipe de France de football a vaincu l'Australie en finale de la Coupe du Monde. L'entraîneur est très satisfait de son équipe et envisage les rencontres futures avec enthousiasme et sérénité.</p>	<p>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.</p>	<p>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.</p>	
<p>Le candidat conservateur a bénéficié du soutien du fort contingent d'évangélistes et de religieux conservateurs qui habitent cet Etat souvent critiqué pour son manque de représentativité à l'échelon national.</p>	<p>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.</p>	<p>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.</p>	<p>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.</p>
	<p>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.</p>	<p>Les Etats-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.</p>	<p>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.</p>
<p>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.</p>	<p>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.</p>	<p>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.</p>	<p>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.</p>


I Pages d'un thème dans la modalité *Sans Titre* pour l'expérience structurée en pêle-mêle

Pages obtenues pour le thème *prix du pétrole*.

Le Journal Mise à jour à 07 h 03 - Paris		Regardez mieux pour mieux comprendre EYELSA
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.	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.	
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.		
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.	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.	
	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.	
		
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.		

Le Journal Mise à jour à 07 h 03 - Paris		Regardez mieux pour mieux comprendre EYELSA
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.	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.	
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.		
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.	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.	
	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.	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.

Le Journal Mise à jour à 07 h 03 - Paris		Regardez mieux pour mieux comprendre EYELSA
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.	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.	
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.	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.	
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.	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.	
	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.	

Le Journal Mise à jour à 07 h 03 - Paris		Regardez mieux pour mieux comprendre EYELSA
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.	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.	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.
	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.	
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.	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.	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.


J Pages d'une liste pour l'expérience structurée en pêle-mêle

Pages obtenues pour tous les thèmes dans une liste, chaque page a une organisation différente.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Le monocoque Garmore du Britannique a démarré dans la troisième étape de l'Around Alone, tour du monde à la voile en solitaire avec escales, entre Auckland et Punta del Este.



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.

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.

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.

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.

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.

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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.


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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.



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.

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.

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.

Aucune preuve scientifique ne permet aujourd'hui de démontrer que l'utilisation de téléphones mobiles présente un risque notable pour la santé. Le communiqué du ministère ne laisse place à aucun doute.

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.

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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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.

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.

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.

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.


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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.



Les transports connaissent 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.

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.

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.

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 maîtrisée.

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.

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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

L'équipe de France de football a vaincu l'Australie en finale de la Coupe du Monde. L'entraîneur est très satisfait de son équipe et envisage les rencontres futures avec enthousiasme et sérénité.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


La plupart des animateurs-moderateurs 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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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.

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 débrasse la plupart des convulsions de l'histoire.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.


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.

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.

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.

Le retour à la vraie vie peut s'avérer douloureux pour ceux qui n'ont pas économisé les dollars pour garnir les matelas qui atténuaient leur atterrissage. C'est le cas de ce basketteur célèbre.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.

Les élections de l'low a 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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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 Chite contre Israël.

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.


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.

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.

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.

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 quelque'un d'important.

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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

Dans le jugement, cité par La Nación, les juges mettent en avant plusieurs arguments de forme, arguant que le type de procédure lancé par le juge Cerda serait inapproprié.


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.

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.

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.

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.

Un groupe d'ingénieurs météorologistes a obtenu le prix Nobel pour ses travaux relatifs à la prévision des événements climatiques à long terme et aux effets de la pollution de l'atmosphère.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.


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.

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.

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.

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.


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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Le chanteur Yannick Noah, ici en concert, a repris sa place à la tête du classement des personnalités préférées des Français et devance Zinedine Zidane.



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.

Le président de l'APM a villendé 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.

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.

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.

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.

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.


Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

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.



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.

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.


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%.

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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.



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.

Il n'est point de bon mercato sans un feuilleton pour l'animer. Les transferts des deux footballeurs n'ont pas électrisés les foules en dehors du premier cercle des admirateurs de l'un et de l'autre.

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.

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.

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.

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.


Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

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.

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.

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.



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.

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.

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.

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.

K Pages d'une liste dans la modalité *Avec Titre* pour l'expérience structurée en mêle-mêle

Pages obtenues pour tous les thèmes dans une liste, chaque page a une organisation différente.

Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Détente
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.

Nature
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.


Conflit
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.

Détente
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.

International
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.

Conflit
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.

Nature
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.



Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Logement
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.


Nature
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.

Social
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 embrasse la plupart des convulsions de l'histoire.

International
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.

Sports
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.

Logement
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.



Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Climat
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.

Justice
Dans le jugement, cité par La Nación, les juges mettent en avant plusieurs arguments de forme, arguant que le type de procédure lancé par le juge Cerda serait inapproprié.


Environnement
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.

International
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.

Social
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.

Environnement
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.

Climat
Un groupe d'ingénieurs météorologistes a obtenu le prix Nobel pour ses travaux relatifs à la prévision des événements climatiques à long terme et aux effets de la pollution de l'atmosphère.



Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Economie
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.

Sports
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.


Finance
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.

Volle
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.

Finance
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.

Economie
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%.

Environnement
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.



Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Football
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.

Sports
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.


Conflit
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.

Culture
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.

Sports
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.

Football
L'équipe de France de football a vaincu l'Australie en finale de la Coupe du Monde. L'entraîneur est très satisfait de son équipe et envisage les rencontres futures avec enthousiasme et sérénité.

Politique
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.



Le Journal

Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Economie
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.

Social
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.


International
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.

Environnement
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.

Culture
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.

Economie
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.

International
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Education
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.

Informatique
La plupart des animateurs-moderateurs 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.


Politique
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.

Nature
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.

Informatique
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 mises en place.

Social
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.

Education
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

International
La Patagonie chilienne est mal connue. Coïncé entre la cordillère des Andes et l'océan Pacifique, elle séduit les amateurs de grands espaces, de solitude et de paysages grandioses.


Justice
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.

Social
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.

Sciences
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.

Sports
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.

Social
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Conflit
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 Chite contre Israël.

International
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.


Social
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.

Médecine
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
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.

Sports
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 quelque'un d'important.

International
Pour l'instant, les télécommunications palestiniennes sous-traitent l'accès à l'Internet à une compagnie privée israélienne. Les réseaux sont au point et on trouve maintenant des cabines téléphoniques publiques à carte.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Justice
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.

International
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.


Sports
Il n'est point de bon mercato sans un feuilleton pour l'animer. Les transferts des deux footballeurs n'ont pas électrisés les foules en dehors du premier cercle des admirateurs de l'un et de l'autre.

Conflit
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.

Nature
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.

International
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.

Conflit
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Education
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.

Economie
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.


Social
Afin d'apprendre une langue étrangère le ministre 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.

Sports
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.

Education
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.

Culture
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.

Social
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Culture
Le chanteur Yannick Noah, ici en concert, a repris sa place à la tête du classement des personnalités préférées des Français et devance Zinédine Zidane.

Environnement
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.


Justice
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.

Social
Bien que les preuves de la discrimination à l'embauche soient difficiles à établir devant les juges, quelques affaires ont pu aboutir à confondre des employeurs et ont entraîné des condamnations.

Economie
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.

Justice
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.


Social
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.



Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Voile
Le monocoque Garmore du Britannique a démarré dans la troisième étape de l'Around Alone, tour du monde à la voile en solitaire avec escales, entre Auckland et Punta del Este.



International
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.

Conflit
Le fondateur de la guerrilla, 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.

Sports
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.

Sciences
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.

Voile
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.

Sports
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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA


Finance
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.

Nature
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.

Economie
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.

Sports
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.

Culture
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.




Economie
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.

Finance
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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Sports
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.



Social
Les transports connaissent 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.

Social
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.

Emploi
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.

Politique
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 maîtrisée.

Culture
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.

Emploi
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.

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

Social
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.

Solidarité
La détresse et la souffrance des réfugiés largement médiatisés, suscitent d'immenses élan de générosité. Des convois d'aide humanitaire constitués par des bénévoles de différentes organisations convergent vers les camps de réfugiés.


Sports
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 prédition. La Française l'a emporté facilement, reléguant sa grande rivale.

Social
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.

Solidarité
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.


International
Les élections de l'Iowa ont été sans doute les plus disputées de l'histoire de cet Etat rural. La participation à ces assemblées a dépassé tous les espoirs prévus par les politiques.

Conflit
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.




Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

<p>Social</p> <p>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.</p>	<p>Politique</p> <p>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.</p>	<p>Sports</p> <p>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.</p> 
<p>Conflit</p> <p>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.</p>	<p>Nature</p> <p>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.</p>	<p>Politique</p> <p>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.</p>
<p>Conflit</p> <p>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.</p>		


Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

<p>Culture</p> <p>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.</p>	<p>Economie</p> <p>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.</p>	<p>Finance</p> <p>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.</p>
<p>Finance</p> <p>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.</p> 	<p>Sports</p> <p>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.</p>	<p>Social</p> <p>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.</p>
	<p>Economie</p> <p>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.</p>	

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

<p>Justice</p> <p>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.</p>	<p>Sports</p> <p>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.</p> 
<p>Astronomie</p> <p>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.</p>	<p>Sciences</p> <p>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.</p>
<p>Sciences</p> <p>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.</p>	<p>Social</p> <p>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.</p>
<p>Astronomie</p> <p>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.</p>	

Le Journal
Mise à jour à 07 h 03 - Paris

Regardez mieux pour mieux comprendre
EYELSA

<p>International</p> <p>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.</p>	<p>Conflit</p> <p>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.</p>
<p>Economie</p> <p>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.</p>	<p>International</p> <p>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.</p>
<p>Sports</p> <p>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.</p> 	<p>Conflit</p> <p>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.</p>
	<p>Social</p> <p>Les enfants jouent au foot avec des balles de tissu 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.</p>

Références

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ : Erlbaum.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036-1060.
- Arani, T., Karwan, M. H., & Drury, C. G. (1984). A variable-memory model of visual search. *Human Factors*, *26*, 631-639.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory : A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation (volume 2)* (p. 89-195). New York : Academic Press.
- Baccino, T., Bellino, C., & Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces*. Paris : Hermès Science Publisher (Lavoisier).
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation : Advances in research and theory*, *8*, 47-89. (New York : Academic Press.)
- Bauer, T., Laham, D., Benz, Z., Dooley, S., Kimmel, J., & Oberbrekling, R. (2005). Text analysis and dimensionality reduction for automatically populating a cognitive model framework. In C. Forsythe, M. Bernard, & T. Goldsmith (Eds.), *Cognitive systems : Human cognitive models in systems design*. Lawrence Erlbaum Associates.
- Bouju, F., & Sperandio, J. C. (1979). *Analyse de l'activité visuelle des contrôleurs d'approche* (Rapport technique N° CO 7911 R 59). Rocquencourt : INRIA.
- Byrne, M. (2005). Cognitive Architectures in HCI : Present Work and Future

- Directions. In *Proceedings of the 11th International Conference on Human Computer Interaction*. Las Vegas, NV.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- Chalon, R., David, B., Beldame, M., Cherief, N., Lasalle, J., & Moinard, J. (2001). L'oculomètre comme support d'évaluation et d'interaction. In A. Blanford, J. Vanderdonckt, & A. Derycke (Eds.), *Conférence conjointe AFHIM-BCS sur l'Interaction Homme-Machine (IHM-HCI)* (p. 117-122).
- Chanceaux, M., Guérin-Dugué, A., Lemaire, B., & Baccino, T. (2008). Towards a model of information seeking by integrating visual, semantic and memory maps. In *Proceedings of the 4th international cognitive vision workshop* (p. 65-78). Lecture Notes in Computer Science 5329, Berlin : Springer Verlag.
- Chanceaux, M., Guérin-Dugué, A., Lemaire, B., & Baccino, T. (2009a). An information search model integrating visual, semantic and memory processes. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society*.
- Chanceaux, M., Guérin-Dugué, A., Lemaire, B., & Baccino, T. (2009b). A model to simulate web users' eye movements. In *Proceedings of the 12th IFIP TC13 Conference in Human-Computer Interaction, INTERACT 2009*. Lecture Notes in Computer Science.
- Chanceaux, M., Guérin-Dugué, A., Lemaire, B., & Baccino, T. (2009c). Simulations cognitives de trajets oculomoteurs lors d'une recherche d'information sur des pages numériques. In *Proceedings of the 21st International Conference of the Association Francophone d'Interaction Homme-Machine*. ACM New York, NY, USA.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshmann, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, *41*, 391-407.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, *18*, 193-222.
- Dickinson, C. A., & Zelinsky, G. J. (2007). Memory for the search path : Evidence for a high-capacity representation of search history. *Vision Research*, *47*, 1745-1755.
- Dix, A. J., Finlay, J. E., Abowd, G. D., & Beale, R. (1998). *Human-computer interaction*. London : Prentice Hall Europe.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433-458.
- Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT : A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*, 777-813.
- Eyrolle, H., Virbel, J., & Lemarié, J. (2008). Effect of incomplete correspondance between document titles and their text on users representations : A cognitive

- and linguistic analysis based on 25 technical documents. *Applied Ergonomics*, *39*, 241-246.
- Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (p. 295-312). Amsterdam : Elsevier.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision : The psychology of looking and seeing* (O. U. Press, Ed.). Oxford.
- Fu, W., & Pirolli, P. L. (2007). SNIF-ACT : a cognitive model of user navigation on the World Wide Web. *Human Computer Interaction*, *22* (4), 355-412.
- Geisler, W. S., & Perry, J. S. (2002). Real-time simulation of arbitrary visual fields. In *Proceedings of the 2002 symposium on eye tracking research & applications* (p. 83-87). New York, NY.
- Glenberg, A. M., & Robertson, D. A. (2000). Grounding Symbols and Computing Meaning : A supplement to Glenberg & Robertson. *Journal of Memory and Language*, *43*, 379-401.
- Halverson, T., & Hornof, A. J. (2007). A minimal model for predicting visual search in human-computer interaction. In *Proceedings of the conference on human factors in computing systems* (p. 431-434).
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*, 219-222.
- Henderson, J. M., Brockmole, J. R., Castelhana, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during search in real-world scenes. In R. P. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements : A window on mind and brain*. Oxford : Elsevier.
- Henderson, J. M., & Smith, T. J. (2009). The influence of clutter on real-world scene search : Evidence from search efficiency and eye movements. *Journal of Vision*, *9*(1), 1-8.
- Hooge, I. T., Over, E. A., van Wezel, R. J., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*, 1901-1908.
- Ho Phuoc, T., Guérin-Dugué, A., & Guyader, N. (2009). A computational saliency model integrating saccade programming. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (p. 57-64).
- Horowitz, T. S. (2006). Revisiting the variable memory model of visual search. *Visual Cognition*, *14*(4-8), 668-684.
- Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets : Remember the targets, forget the search. *Perception and Psychophysics*, *63*(2), 272-285.
- Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, *13*, 1304-1318.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489-1506.

- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual-attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 1254-1259.
- John, B. E., & Salvucci, D. D. (2005). Multi-purpose prototypes for assessing user interfaces in pervasive computing systems. *IEEE Pervasive Computing*, 4(4), 27-34.
- Katsanos, C., Tselios, N., & Avouris, N. (2009). Are 10 users enough for measuring information scent? In S. V. LNCS (Ed.), *Proceedings of the 12th IFIP TC13 Conference in Human-Computer Interaction, INTERACT 2009*.
- Kaur, I., & Hornof, A. J. (2005). A comparison of LSA, WordNet and PMI-IR for predicting user click behavior. In *Proceedings of ACM CHI 2005 : Conference on human factors in computing systems* (p. 51-60). New York : ACM.
- Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, 4(12), 391-438.
- Kieras, D. E., & Polson, P. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22(4), 365-394.
- King, J. W., & Kutas, M. (1995). A brain potential whose latency indexes the length and frequency of words. *CRL Newsletter*, 10(2).
- Kintsch, W. (1998). *Comprehension : A paradigm for cognition*. Cambridge, U. K. and New York : Cambridge University Press.
- Kintsch, W. (2000). Metaphor comprehension : A computational theory. *Psychonomic Bulletin & Review*, 7(2), 257-266.
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A comprehension-based model of web navigation and its application to web usability analysis. In *People and Computers XIV - Usability or Else! (Proceedings of HCI 2000)* (p. 357-373).
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2005). Cognitive architecture for website design and usability evaluation : Comprehension and information scent in performing by exploration. In *Proceedings of HCI-International 2005*.
- Kitajima, M., Polson, P. G., & Blackmon, M. H. (2007). CoLiDeS and SNIF-ACT : Complementary Models for Searching and Sensemaking on the Web. In *Human Computer Interaction Consortium (HCIC) 2007 Winter Workshop*.
- Klein, R. M., & MacInnes, J. W. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, 10(4), 346-352.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention : towards the underlying neural circuitry. *Human Neurobiology*, 4, 219-27.
- Landauer, & Dumais, S. (1997). A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104, 211-240.

- Landauer, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- Landauer, D. C., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates.
- Landragin, F. (2004). Saillance physique et saillance cognitive. *Cognition, Représentation, Langage (CORELA)*, 2(2).
- Lemaire, B. (1998). Models of high-dimensional semantic spaces. In *Proceedings of the 4th Int. Workshop on Multistrategy Learning (MSL'98)*. Desenzano.
- Lemaire, B., Denhière, G., Bellissens, C., & Jhean-Larose, S. (2006). A computational model for simulating text comprehension. *Behavior Research Methods*, 38(4), 628-637.
- Lemaire, B., Mandin, S., Dessus, P., & Denhière, G. (2005). Computational cognitive models of summarization assessment skills. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society (CogSci' 2005)* (p. 1266-1271).
- Long, M. (2000). Literacy. In M. Long (Ed.), *The psychology of education* (p. 215-243). Taylor & Francis Ltd.
- MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2000). A model of human performance on the travelling salesperson problem. *Memory & Cognition*, 28, 1183-1190.
- Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press.
- Martin, D. L., & Berry, M. W. (2007). Mathematical Foundations Behind Latent Semantic Analysis. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading : Frequency of refixating a word. *Perception and Psychophysics*, 46, 245-53.
- Miller, G. A. (1956). The magical number seven, plus or minus two : Some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97.
- Miller, G. A. (1995). WordNet : A lexical database for english. *Communications of the ACM*, 38(11), 39-41.
- Moores, L. E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, 6(2), 8.
- Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, 38(12), 1805-1815.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2), 205-231.
- Navarro, G. (2001). A guided tour to approximate string matching. *Computing Surveys*, 33(1), 31-88.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA : Harvard University Press.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene : The role of global

- image features in recognition. *Progress in Brain Research : Visual perception*, 155, 23-36.
- Pirolli, P., & Fu, W. (2003). SNIF-ACT : A Model of Information Foraging on the World Wide Web. In *9th International Conference on User Modeling* (p. 45-54).
- Pomplun, M. (2007). Advancing area activation towards a general model of eye movements in visual search. In W. Gray (Ed.), *Integrated models of cognitive systems*. New York : Oxford.
- Pomplun, M., & Ritter, H. (1999). A three-level model of comparative visual search. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty first annual conference of the cognitive science society* (p. 543-548).
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (p. 531-556). London : Erlbaum.
- Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, 331, 163-166.
- Rehder, B., Schreiner, M. E., Wolfe, M. B. W., Laham, D., Landauer, T. K., & Kintsch, W. (1998). Using Latent Semantic Analysis to assess knowledge : Some technical considerations. *Discourse Processes*, 25(2/3), 337-354.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125-157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading : Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445-526.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2006). E-Z Reader : A cognitive-control, serial-attention model of eye-movement control during reading. *Cognitive Systems Research*, 7, 4-22.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 445-471.
- Ritter, F. E., & Young, R. M. (2001). Embodied models as simulated users : introduction to this special issue on using cognitive models to improve interface design. *International Journal of Human-Computer Studies*, 55, 1-14.
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion, a measure of display clutter. In *SIGCHI 2005* (p. 760-770).
- Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, 1, 201-220.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*. Manchester, UK.
- Steinhauer, S. R. (2002). *Cognition, psychopathology, and recent pupil studies* (Report technique). University of Pittsburgh School of Medicine.
- Sun, R. (2008). Introduction to computational cognitive modeling. In *The cam-*

- bridge handbook of computational psychology* (p. 3-19). New York : Cambridge University Press.
- Takeda, Y. (2004). Search for multiple targets : Evidence for memory-based control of attention. *Psychonomic Bulletin and Review*, 11, 71-76.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection : effects of scale and time. *Vision Research*, 45(5), 643-659.
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2) :5, 1-18.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.
- Tollinger, I., Lewis, R. L., McCurdy, M., Tollinger, P., Vera, A., & Howes, A. (2005). Supporting efficient development of cognitive models at multiple skill levels : Exploring recent advances in constraint-based modeling. In *Proceedings of the Conference on Human Factors in Computing Systems* (p. 411-420).
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes : The role of global features in object search. *Psychological Review*, 113(4), 766-786.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136.
- Turney, P. (2001). Mining the web for synonyms : PMI-IR versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning* (p. 491-502).
- West, R. L., & Emond, B. (2001). Can cognitive modeling improve usability testing and rapid prototyping? In *Proceedings of the Fourth International Conference on Cognitive Modeling* (p. 271-273). Mahwah, NJ : Lawrence Erlbaum Associates.
- Wolfe, J. M. (2007). Guided search 4.0 : Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (p. 99-119). New York : Oxford.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York : Plenum Press.

Résumé

Ce travail de thèse étudie la combinaison des processus visuels et sémantiques dans le traitement d'interfaces textuelles, comme des pages web. La méthodologie employée est la simulation de modèles cognitifs, une approche visant à concevoir un programme informatique s'inspirant de modèles cognitifs théoriques et reproduisant au mieux le comportement humain. Notre modèle simule les trajets oculomoteurs d'un utilisateur moyen lors d'une recherche d'information. Les processus mis en jeu dans ce type de tâche sont modélisés afin de reproduire les mouvements oculaires humains enregistrés lors de différentes expérimentations. Aux modèles des processus visuels et sémantiques est ajouté un modèle des processus mnésiques sous-jacents à la recherche d'information. Pour la partie visuelle, le modèle s'inspire des cartes de saillance qui prédisent les zones de l'écran susceptibles d'attirer l'attention, en fonction des informations à bas niveau (couleur, orientation et contraste), ainsi que des propriétés physiologiques de l'œil humain. Pour la partie sémantique, la technique de mesure de similarités sémantiques entre le but de l'utilisateur et les différentes parties de la page employée est LSA (Latent Semantic Analysis) (Landauer et al., 1998). Pour la partie mnésique, le mécanisme d'inhibition de retour (Klein & MacInnes, 1999) et le Variable Memory Model (Horowitz, 2006) sont utilisés. Le travail de thèse intègre trois parties : conception d'un modèle théorique d'interaction, conception de l'outil de simulation, et mise en place d'expérimentations psychophysiques de relevés oculométriques permettant de valider et d'affiner le modèle proposé.

Abstract

This thesis examines the combination of visual and semantic processes in information seeking task on textual interfaces such as web pages. The methodology used is the simulation of cognitive models. This approach aims to design a program based on theoretical cognitive models replicating human behaviour. Our model simulates the oculomotor scan path of an average user searching for information. The processes involved in this kind of tasks are modelled to replicate human eye movements recorded during different experiments. Models of visual and semantics processes are added to a model of memory processes inherent in information retrieval task. For the visual part, we go 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. For the semantic part, the method used to measure semantic similarities between the search goal of the user and the different parts of the page is LSA (Latent Semantic Analysis) (Landauer et al., 1998). For the memory part the mechanism of inhibition of return (Klein & MacInnes, 1999) and the Variable Memory Model (Horowitz, 2006) are used. The thesis includes three parts : designing a theoretical model of interaction, designing a simulation tool, and developing psychophysical experiments with eye-tracking techniques to validate and refine the proposed model.