

# Simulations cognitives de trajets oculomoteurs lors d'une recherche d'information

*Myriam Chanceaux*

Laboratoire TIMC-IMAG, Université de Grenoble  
Domaine de la Merci  
38700 La Tronche, FRANCE  
myriam.chanceaux@imag.fr

## RESUME

Nous présentons ici un modèle informatique qui simule le parcours du regard d'un utilisateur qui recherche une information sur des pages Web. Les mouvements oculaires sont guidés par le besoin d'information, par les caractéristiques visuelles des stimuli et par ce qui a déjà été traité et mémorisé. Notre modèle tient donc compte à la fois des informations sémantiques (top-down) et visuelles (bottom-up) et intègre un modèle de mémoire afin de prédire la direction de l'attention. Ce modèle fonctionne au niveau du paragraphe, mais aussi au niveau du mot. Pour valider ce modèle, plusieurs expérimentations avec oculomètre ont été réalisées dans lesquelles chaque participant était mis en contexte de recherche d'information sur une page d'un pseudo journal d'information en ligne. Le modèle réalisé prédit le parcours des fixations oculaires durant la scrutation de la page et on observe une bonne correspondance entre les trajets oculaires empiriques et ceux prédits par le modèle.

**MOTS CLES :** Modélisation Cognitive, Mouvements Oculaires, Mémoire, Sémantique, Visuel, Pages Web, Recherche d'Information

## ABSTRACT

We present here a computational model that simulates human movements of a web user during an information seeking task. Eye movements are guided by the need for information, by the visual characteristics of the stimuli and by what has already been processed and stored. Our model takes into account both semantic (top-down) and visual (bottom-up) information and includes a memory model to predict direction of attention. This model operates at the text block level, and at the word level. To validate this model, we asked participants to search information in a

pseudo online newspaper, and we compared their scanpaths with those of the model. We observed a good correspondence between simulation and empirical observation.

**CATEGORIES AND SUBJECT DESCRIPTORS:** H.5.2 User Interfaces: Evaluation/methodology

**GENERAL TERMS:** Human Factors, Experimentation

**KEYWORDS:** Cognitive Model, Memory, Semantics, Vision, Web pages, Eye Movements, Information Search

## INTRODUCTION

Le but de cette thèse est de proposer un modèle cognitif de la navigation dans un document structuré, visant à faciliter la caractérisation de son utilisabilité. Ce type de modèle prend 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. Cette méthodologie possède une longue histoire, les modèles plus récents étant par exemple CoLiDeS [6] ou SNIF-ACT [10]. 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).

La tâche que nous cherchons à modéliser et à simuler est la recherche d'informations dans un document structuré, idéalement une page Web (même si pour le moment nous n'envisageons pas toute la complexité d'une telle interface). Un tel document peut être vu comme la réunion de deux types d'entités : des unités textuelles multiples (à l'opposé donc 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 : (1) la modélisation de la compréhension de textes, depuis les modèles de lecture jusqu'aux modèles de traitement sémantique [7], qui sont habituellement appliqués à des textes uniques et (2) la modélisation des processus visuels [5], traditionnellement appliqués à des scènes visuelles sans contenus textuels. Les interactions entre les processus visuels et les processus sémantiques constituent donc

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IHM 2009, 13-16 Octobre 2009, Grenoble, France

Copyright 2009 ACM 978-1-60558-461-4/09/10 ...\$5.00.

la spécificité de cette thèse.

La tâche que nous souhaitons étudier concerne la recherche d'informations et non la simple navigation sans but. Nous cherchons donc à simuler un humain ayant un objectif de recherche précis. Notre objectif est aussi de modéliser le comportement humain de navigation à un niveau de précision suffisamment fin. Cela nous oblige à nous situer au niveau des fixations et des saccades oculaires. Pour résumer, notre but est donc de prédire le trajet oculaire moyen d'utilisateurs confrontés à un document structuré donné, avec une intention précise (c'est à dire un besoin d'information défini par la consigne).

## TACHE

La tâche consiste à rechercher une information dans une pseudo page Web, ressemblant à un journal en ligne. Le but est de trouver le bloc de texte qui correspond le mieux à l'expression donnée. Voici quelques-uns des 20 thèmes : *réchauffement climatique, hausse de la bourse, victoire des footballeurs, réforme de l'enseignement....*

Chaque page correspond à un thème et contient 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 image. Les textes sont divisés en 3 catégories : 2 textes sont fortement associés au thème et sont donc des cibles potentielles (AF) ; 2 textes ont un rapport sémantique, mais faible, avec ce thème (Afa) ; et 3 textes n'ont aucune association avec le thème (AA).

Dans la première expérience les blocs sont alignés sur deux colonnes et peuvent avoir un fond blanc ou de couleur. Il y a trois couleurs par image, qui ont été choisies de manière à avoir le même contraste avec le texte écrit en noir, avec le fond blanc et entre elles. Dans chaque page, un bloc AF, un Afa et un AA ont un fond en couleur. Ces 3 blocs sont choisis aléatoirement uniformément. Dans la seconde expérience les blocs ont des tailles différentes et ne sont plus strictement alignés. De plus, un titre a été attribué à chaque paragraphe, comme on en trouve couramment dans les pages Web. La figure 1 montre un exemple de page pour l'expérimentation 2.

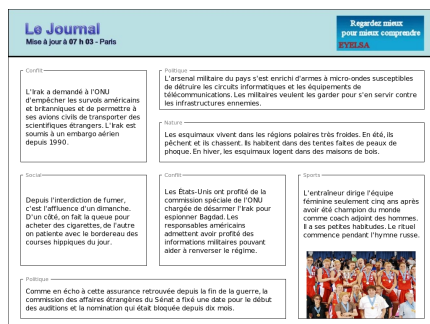


Figure 1 : Exemple de page pour l'expérimentation 2, le thème est *conflit en Irak*.

## MODELE

Etant 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, qui sont des coordonnées XY. Nous avons considéré qu'il y avait deux niveaux distincts de navigation dans la page. Le premier consiste à aller de blocs en blocs (niveau du bloc), et le second opère à l'intérieur de chaque bloc (niveau du mot).

## Au niveau du bloc

Au niveau du bloc, une analyse de trajets humains nous a conduit à observer deux temps dans le parcours oculaire. Tout d'abord une stratégie globale 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) et physiologiques (les grandes saccades sont coûteuses par exemple). Nous avons repris ces paramètres dans la modélisation en pondérant les directions. Le modèle a donc des paramètres  $\alpha_H$  et  $\alpha_G$  qui correspondent respectivement au poids d'un déplacement vers le haut, et au poids d'un déplacement vers la gauche. Ces paramètres entraînent un parcours de blocs en blocs, comme le montre la figure 2. L'algorithme que nous avons implémenté se base sur celui de la résolution du problème du voyageur du commerce le plus cognitivement plausible : l'algorithme glouton par insertion du noeud le moins coûteux [8]. Le coût étant dépendant de la distance entre blocs et de la direction de la transition. Les données empiriques nous permettront d'estimer les valeurs de  $\alpha_H$  et  $\alpha_G$  les plus appropriées.

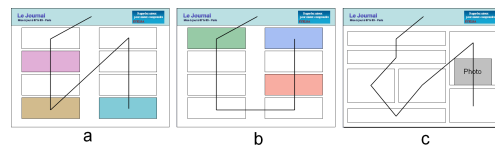


Figure 2 : (a) Trajet en N inversé généré par le modèle, avec les paramètres  $\alpha_H=1.5$  et  $\alpha_G=2$ , (b) Trajet en U généré par le modèle, avec les paramètres  $\alpha_H=1$  et  $\alpha_G=2$  et (c) Trajet généré par le modèle avec les paramètres  $\alpha_H=1.2$  et  $\alpha_G=1.3$ .

Une fois tous les blocs visités, une seconde partie du parcours concerne les revisites des blocs considérés intéressants. Les revisites sont basées sur trois processus cognitifs: visuel, sémantique et mnésique. Ces processus peuvent être en conflit : le processus visuel aura tendance à privilégier les blocs proches ou visuellement saillants, le processus de la mémoire tendra vers ceux qui ont été vus en premier (et donc oubliés) et le processus sémantique voudra éviter les blocs les moins associés au thème de la recherche. Notre modèle inclut un mécanisme pour intégrer ces trois sources d'information.

La valeur visuelle  $P_V$  de chaque bloc  $B_i$  tient compte de sa proximité géographique avec le bloc fixé suivant la courbe classique de l'acuité visuelle en fonction de l'excentricité [4]. Au niveau de la saillance visuelle, étant donné la simplicité des stimuli, un simple biais est ajouté

afin de légèrement favoriser les blocs dont le fond est coloré.

Les niveaux sémantique et mnésique sont intimement liés. Chaque bloc  $B_i$  reçoit un poids  $P_{MS}$  qui correspond à son attractivité. Lorsque le modèle quitte un bloc, il lui attribue un poids correspondant à la similarité sémantique entre le thème et la partie du bloc qui a été traitée. Par la suite ce poids va avoir tendance à augmenter, pour favoriser les blocs les plus anciens et simuler ainsi un mécanisme d'oubli.

A 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  $B_i$ :

$$P(B_i) = \alpha_{MS} \cdot P_{MS}(B_i) + \alpha_V \cdot P_V(B_i)$$

Les paramètres  $\alpha_{MS}$  et  $\alpha_V$  ont été déterminés grâce à une expérimentation préalable [1]. La mémoire associée au sémantique a un rôle plus important  $\alpha_{MS} = 0.65$  et  $\alpha_V = 0.35$  Lorsque un bloc est sélectionné, la seconde partie du modèle prend le relais à l'intérieur du bloc.

### Au niveau du mot

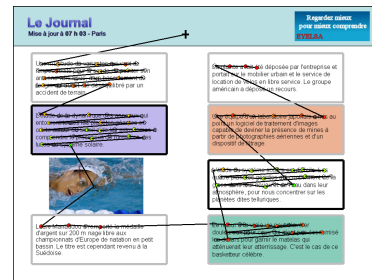
A ce niveau lexical, le modèle utilise comme les humains l'information sémantique contenue dans la séquence de mots fixés et sa similarité avec le thème recherché. Une telle tâche de recherche d'informations se distingue d'une tâche de lecture dans laquelle le paragraphe est lu intégralement. Les humains n'hésitent pas à abandonner le traitement d'un bloc dès qu'ils ont acquis suffisamment d'informations sur son intérêt. C'est cette stratégie que nous avons essayé de modéliser : le modèle parcourt les mots (hors mots-outils) les uns après les autres en décidant après chacun d'eux de quitter le bloc ou non en calculant les mesures de similarités LSA (*Latent Semantic Analysis, indexation sémantique latente*) [7, 2]. LSA se base sur un corpus de plusieurs millions de mots (ici, tous les articles de l'année 1999 du quotidien *Le Monde*) à partir duquel un espace sémantique de 300 dimensions est généré. Dans cet espace tout document (mot ou groupe de mots) est représenté par un vecteur. La similarité entre deux documents correspond au cosinus entre leurs vecteurs.

Lorsque les paragraphes sont associés à un titre (expérience 2) le modèle a un comportement particulier. En effet le titre a une valeur sémantique différente des autres mots du paragraphe étant donné qu'il est représentatif du texte. Un titre en lien avec le thème de la recherche va inciter à la lecture et au contraire un titre sans rapport avec le thème va entraîner une transition vers un autre bloc. Le modèle prend en compte cette similarité entre le titre et le thème pour décider de la prochaine fixation.

Notre modèle n'analyse pas la photo, pas plus qu'il ne tient compte de sa pertinence visuelle. Par conséquent, il considère qu'il s'agit d'un bloc sans texte et ne fait qu'une fixation sur le milieu de la photo. La figure 3 présente un exemple de scanpath sur l'image de l'expérimentation 1 présentée plus tôt, avec comme paramètres  $\alpha_H=1$  et

$\alpha_G=2$ .

Les entrées du modèle sont les coordonnées XY de chaque mot et de chaque paragraphe.



**Figure 3 :** Exemple d'un scanpath généré par le modèle. Les blocs abandonnés en raison de leur forte similarité avec le thème sont entourés en noir. Les blocs abandonnés en raison de leur faible similarité au thème sont entourés en gris.

### TEST DU MODELE

Afin de tester la validité de notre modèle, nous avons enregistré les trajets oculomoteurs de participants et comparé leur comportement à celui du modèle.

### Expériences

Les mouvements oculomoteurs de 38 participants ont été enregistrés avec un oculomètre SR EyeLink II pour l'expérience 1 et ceux de 15 participants avec l'EyeLink1000 pour la seconde. Chaque essai commence par une instruction suivie d'une croix de fixation. Les participants doivent trouver le meilleur bloc en fonction du thème, sans délai maximum, par exemple trouver le bloc le plus proche du thème *Chef de la Russie*.

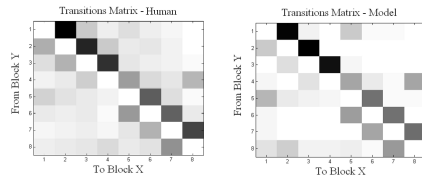
### Ajustement du modèle

Comme nous l'avons vu précédemment, deux paramètres ( $\alpha_H$  et  $\alpha_G$ ) modulent la stratégie globale du modèle et donc la forme du trajet. A partir de distance de Levenshtein [9] entre les trajets des sujets et ceux du modèle de l'expérience 1, nous avons fixé des intervalles pour ces 2 paramètres afin de reproduire la diversité des trajets empiriques.

Pour chaque essai, 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.

### Comparaison entre le modèle et les humains

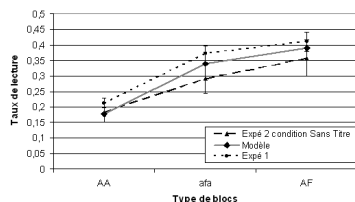
**Au niveau du bloc** Des matrices de transitions (figure 4) entre chacun des 8 blocs nous montrent les similitudes entre les trajets des humains et ceux du modèle. La couleur de la case (i,j) dénote le nombre de transitions du bloc i au bloc j. Plus la couleur est foncée, plus le nombre de transitions est important. 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 humains.



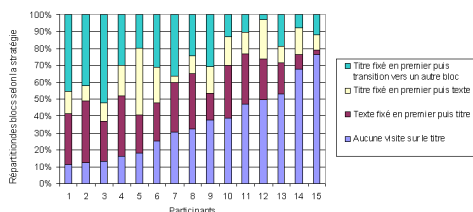
**Figure 4 :** Matrice de transitions entre les blocs pour les participants et pour le modèle sur l'expérimentation 1.

Pour la seconde partie du parcours nous avons regardé quels étaient les blocs refixés. La prise en compte des informations sémantiques, visuelles et mnésiques dans le modèle nous donne des résultats proches de ceux des humains avec dans les deux cas plus de revisites pour les blocs intéressants que pour les autres.

**Au niveau du mot** À ce niveau, nous nous sommes penchés sur le nombre de fixations dans un bloc selon la catégorie du bloc (AF, Afa ou AA) lors de la première visite du bloc. Les résultats montrent que les blocs AA sont quittés plus tôt que les autres (moins de fixations sur ces blocs) et que les blocs AF sont les plus lus (voir figure 5). Pour les blocs avec le titre la stratégie n'est pas la même, et nous avons voulu observer plus en détail le comportement des participants. Nous avons donc classifié les blocs en 4 catégories (voir figure 6). Cette répartition nous a montré une grande variabilité inter individuelle. La version future du modèle essaie de rendre compte de ces différentes stratégies.



**Figure 5 :** Taux de lecture (nombre de fixations divisé par nombre de mots) pour les participants et le modèle selon le type de bloc (association forte, association faible, aucune association).



**Figure 6 :** Répartition des blocs par rapport à la visite ou non du titre et ses conséquences par sujet.

## DISCUSSION

L'évaluation des interfaces est un processus coûteux qui peut être accéléré par le recours à une simulation fondée sur un modèle cognitif plausible de l'utilisateur. Ce modèle peut être amené à évoluer pour prendre en compte

les caractéristiques avancées des interfaces, notamment le côté interactif, mais aussi les images en incluant notamment un véritable modèle de saillance visuelle. Il pourrait être alors complémentaire d'outil automatique basé sur des règles pour un "bon design de page Web"[3].

A terme, on peut par exemple imaginer donner au modèle une page Web nouvellement conçue et un objectif sous la forme d'une séquence de mots et observer le trajet oculaire probable d'un utilisateur afin d'en déduire les éléments d'interface perturbateurs ou facilitateurs pour la réalisation de cet objectif.

## BIBLIOGRAPHIE

1. Chanceaux, M., Guérin-Dugué, A., Lemaire, B., and Baccino, T. Towards a model of information seeking by integrating visual, semantic and memory maps. In *Proceedings of the 4th International Cognitive Vision Workshop*, volume LNCS 5329, pages 65–78, 2008.
2. Deerwester, S., Dumais, S., Furnas, G. W., Landauer, T. K., and Harshman, R. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407, 1990.
3. Faraday, P. Visually critiquing web pages. In *Proceedings of the 6th Conference on Human Factors & the Web*, 2000.
4. Geisler, W. S., and Perry, J. S. Real-time simulation of arbitrary visual fields. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 83–87, New York, NY, 2002.
5. Itti, L., and Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.
6. Kitajima, M., Blackmon, M. H., and Polson, P. G. 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)*, pages 357–373, 2000.
7. Landauer, T., McNamara, D., Dennis, S., and Kintsch, W. *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates, 2007.
8. MacGregor, J. N., Ormerod, T. C., and Chronicle, E. P. A model of human performance on the travelling salesperson problem. *Memory & Cognition*, 28:1183–1190, 2000.
9. Navarro, G. A guided tour to approximate string matching. *Computing Surveys*, 33(1):31–88, 2001.
10. Pirolli, P., and Fu, W. SNIF-ACT: A model of information foraging on the world wide web. In *9th International Conference on User Modeling*, volume 2702 of *Lecture Notes in Artificial Intelligence*, pages 45–54, 2003.