

Clément Arsenault\*

Élaine Ménard

École de bibliothéconomie et des sciences de l'information, Université de Montréal,  
C.P. 6128, succ. Centre-ville, Montréal (QC) H3C 3J7

## Recherche des titres débutant par un article : analyse du comportement de recherche

**Résumé :** Cette étude porte sur le comportement des usagers lors du repérage par titre dans les catalogues de bibliothèque. Nous observons si les problèmes causés par la présence des articles initiaux ou de mots homographes à ces articles dans certains titres affectent le repérage en termes de succès et d'effort requis.

**Abstract:** This study examines user behaviour during know-item retrieval using title index in library catalogues. Our observations concentrate on the problems caused by the presence of an initial article or of a word homograph to an article. Measures of success and effectiveness are taken to determine if retrieval is affected in such cases.

### 1. Énoncé du problème

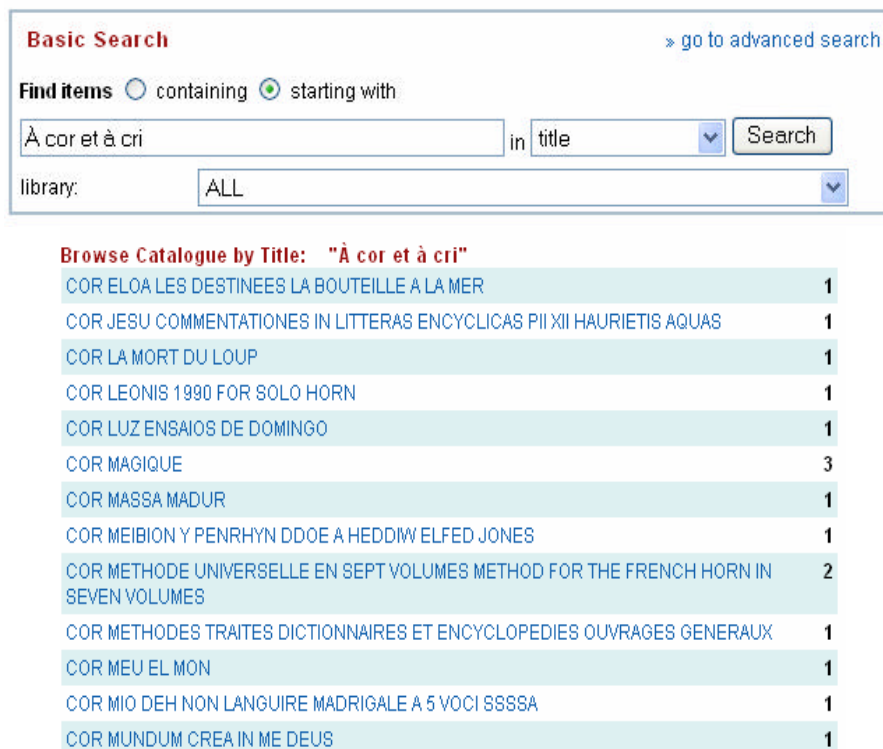
Une première étude de cas effectuée sur le catalogue Atrium de l'Université de Montréal visait à étudier les problèmes de repérage causés pas la présence d'articles, ou encore de mots homographes à des articles, au début des titres de monographies. Cette étude a déjà permis d'identifier les articles initiaux causant les problèmes, et d'évaluer la proportion de notices affectées par ces problèmes. Entre autres choses, l'étude révélait premièrement que l'article anglais « *a* » était responsable à lui seul du deux-tiers des cas problèmes (due à l'abondance de titres débutant par un homographe, p. ex. *À propos de...*) et deuxièmement qu'une proportion non négligeable de notices étaient affectées pas ces problèmes et devenaient par le fait même potentiellement plus difficiles à repérer (Arsenault et Ménard 2005, 7). Toutefois, cette première étude ne nous renseignait pas sur les comportements de recherche des usagers face à ces problèmes de repérage. Par conséquent, il semblait intéressant de poursuivre notre étude et d'examiner l'étendue des problèmes de repérage des titres affectés, du point de vue de l'utilisateur. Une telle étude permettrait d'identifier concrètement de quelle manière l'utilisateur effectue ses recherches dans l'index des titres des catalogues de bibliothèques lorsque les titres débutent par un article ou un mot homographe. En outre, il serait possible dans cette étude de mesurer le taux de succès de l'utilisateur lors du repérage des notices bibliographiques difficilement accessibles par la présence d'un des ces mots en début de champ. Il serait également possible d'établir si le problème en cause a un impact sur le niveau d'efficacité du repérage des monographies (temps et effort requis pour effectuer une tâche de repérage) lors de l'utilisation de l'index de titres et de mesurer empiriquement la portée de ce problème de repérage en situation réelle. Ces informations seraient fort utiles pour améliorer l'efficacité et le taux de succès du repérage dans les

---

\* Cette étude est gracieusement subventionnée par l'Association pour l'avancement des sciences et techniques de la documentation (ASTED). Nous tenons également à remercier l'équipe de la bibliothèque du Collège de Maisonneuve pour nous avoir fourni accès à leurs locaux, ainsi que M<sup>me</sup> Carole Paradis du Bureau des systèmes du réseau des bibliothèques de l'Université de Montréal pour nous avoir fourni les journaux de transaction du catalogue Atrium.

systèmes d'information à contenu multilingue, tels les catalogues, tant au niveau de la formation des usagers que du design et de la conception des systèmes de repérage.

Rappelons un peu le problème. Lorsqu'un usager effectue une recherche dans un catalogue de bibliothèques, il arrive que des algorithmes soient intégrés au module de repérage afin de détecter automatiquement la présence inopportune d'un article initial dans la requête. Étant donné que les articles initiaux sont, dans la grande majorité des cas, éliminés des entrées lors de l'indexation des titres, si l'utilisateur conserve un article initial dans sa requête, l'algorithme se charge alors d'éliminer ce mot afin de positionner l'utilisateur au bon endroit dans l'index, ce qui peut être, dans certains cas, fort utile. Toutefois, dans la grande majorité des cas, ces algorithmes ne sont pas suffisamment intelligents pour détecter certaines subtilités linguistiques, ce qui résulte quelquefois en un problème au moment du repérage. Ainsi, il arrive que les mots homographes à des articles soient erronément retranchés des requêtes. Par exemple, lorsque l'utilisateur tente de retrouver le titre *À cor et à cri*, il se retrouvera à la mauvaise position dans l'index de titres (à la lettre « C » alors que le document a été indexé sous la lettre « A ») car lors de cette recherche, l'algorithme utilisé par le catalogue interprète le « À » comme l'article indéfini anglais « a » et retranche automatiquement ce mot de la requête. L'utilisateur se retrouvera au mauvais endroit dans l'index et aura donc la fausse impression que ce document ne se trouve pas en collection (la notice se trouve bel et bien dans le catalogue mais à la lettre « A »), comme le démontre l'exemple de la figure 1 :



**Basic Search** [» go to advanced search](#)

Find items  containing  starting with

À cor et à cri in title Search

library: ALL

**Browse Catalogue by Title: "À cor et à cri"**

COR ELOA LES DESTINEES LA BOUTEILLE A LA MER	1
COR JESU COMMENTATIONES IN LITTERAS ENCYCLICAS PII XII HAURIETIS AQUAS	1
COR LA MORT DU LOUP	1
COR LEONIS 1990 FOR SOLO HORN	1
COR LUZ ENSAIOS DE DOMINGO	1
COR MAGIQUE	3
COR MASSA MADUR	1
COR MEIBION Y PENRHYN DDOE A HEDDIW ELFED JONES	1
COR METHODE UNIVERSELLE EN SEPT VOLUMES METHOD FOR THE FRENCH HORN IN SEVEN VOLUMES	2
COR METHODES TRAITES DICTIONNAIRES ET ENCYCLOPEDIES OUVRAGES GENERAUX	1
COR MEU EL MON	1
COR MIO DEH NON LANGUIRE MADRIGALE A 5 VOCI SSSSA	1
COR MUNDUM CREA IN ME DEUS	1

**Figure 1 — Exemple de recherche infructueuse en mode « parcourir index »**

Dans le cas précis de ce catalogue, l'utilisateur doit choisir entre deux options de recherche : soit la recherche par le mode « mots-clés » (*containing*), soit la recherche par le mode « parcourir index » (*starting with*). S'il choisit la seconde option (*starting with*), l'utilisateur pourrait facilement conclure que le document recherché ne se trouve pas dans le catalogue puisqu'il ne peut visualiser l'index au bon endroit. S'il choisit la première option (*containing*), il se voit confronté à un autre type de difficultés. En effet, dans le

catalogue de l'Université de Toronto (en avril 2006), en utilisant la requête « cor » l'utilisateur obtient 1160 résultats qu'il devra laborieusement parcourir pour retrouver le document voulu ; la requête « cri » quant à elle mènera à un ensemble de 219 notices, un résultat déjà plus tolérable mais néanmoins élevé si on considère que le but ici est de ne retracer qu'un titre en particulier. Mentionnons toutefois que la requête « cor cri » (l'opérateur booléen AND étant implicite) donne 5 résultats ce qui est très acceptable. Par contre, il arrive que certains titres n'offrent qu'un choix très restreint de termes (p. ex. *À la française* ou encore *À tous*) ce qui mène indubitablement à des ensembles très larges, 6 608 et 2 093 titres respectivement pour une recherche de ces titres en mode « mots-clés ».

Aussi pouvons-nous nous demander de quelle manière l'utilisateur effectue habituellement ses recherches dans les catalogues de bibliothèques, s'il parvient à retrouver les documents recherchés mais surtout, s'il aura tendance à modifier ses habitudes en cas d'insuccès. En effet, dans l'éventualité où l'utilisateur voit que sa requête ne repêche aucun document il peut décider de poursuivre en formulant une autre requête ou encore en utilisant un autre mode de recherche, lorsque disponible. Toutefois, il pourrait également changer de catalogue ou simplement mettre fin à sa recherche.

## **2. Objectifs de l'étude**

Le but de la présente étude est d'étudier l'étendue des problèmes de repérage susmentionnés du point de vue de l'utilisateur. Afin de réaliser ce but, trois objectifs majeurs ont été définis :

1. Identifier selon quel mode de recherche l'utilisateur effectue ses recherches dans l'index de titres des catalogues de bibliothèques lorsque les titres débutent par un article ;
2. Vérifier si la présence d'un mot homographe à un article erronément considéré comme un article initial affecte le taux de succès des usagers (la proportion des notices repérées) lors du repérage dans l'index de titres ;
3. Établir si le problème identifié a un impact sur le niveau d'efficacité du repérage des monographies (temps et effort requis pour effectuer une tâche de repérage) lors de l'utilisation de l'index de titres.

La réalisation de ces trois objectifs permettra de mesurer empiriquement la portée de ce problème de repérage. Ces données seront essentielles pour comprendre le phénomène du repérage des titres et du classement des articles initiaux dans les systèmes de repérage d'information automatisés. La revue de la littérature effectuée lors de la préparation de ce projet démontre que ce problème est peu étudié et qu'il est important de s'attarder plus à fond sur ce phénomène (Bourne 1975 ; Nielsen & Pyle 1995 ; Seeman 2002 ; Kan & Poo 2005). La recherche par titre constitue encore le type de recherche le plus fréquent dans les catalogues de bibliothèques (Larson 1991). Il est donc souhaitable qu'il soit le plus efficace possible.

Nous avons utilisé une méthode expérimentale contrôlée pour effectuer notre collecte de données ce qui a permis de mesurer l'étendue du problème dans un catalogue spécifique. Il sera éventuellement possible d'étendre la recherche à d'autres catalogues, d'autres environnements, ou bien d'en mesurer l'impact direct sur l'utilisateur en milieu de recherche réel. Les résultats de l'étude pourront être utilisés dans le développement d'algorithmes

de repérage plus efficaces afin de mieux paramétrer la recherche par titre dans les systèmes d'information multilingues. Comme les catalogues de bibliothèques sont en quelque sorte la porte d'entrée des collections documentaires, il est essentiel de configurer les systèmes pour maximiser l'efficacité de repérage et le taux de succès, et par conséquent accroître la satisfaction des usagers.

### 3. Méthodologie

Étant donné le caractère exploratoire de cette étude et des moyens limités mis à notre disposition, nous avons choisi de réaliser une étude de cas avec le catalogue de l'Université de Toronto <<http://main.library.utoronto.ca/>>. Ce catalogue a été sélectionné en raison du fait qu'un algorithme de détection a été intégré au module de repérage pour détecter la présence des trois articles anglais «*a*», «*an*» et «*the*». Des journaux de transactions de requêtes effectuées dans le catalogue Atrium (Université de Montréal) ont également été étudiés.

Afin de rencontrer les trois objectifs susmentionnés, les quatre questions de recherche suivantes ont été formulées :

- Question 1 : Habituellement, l'utilisateur conserve-t-il ou omet-il les articles initiaux de ses requêtes lors de ses recherches en mode « parcourir index » dans l'index de titres ?
- Question 2 : Pour effectuer des recherches de titres connus, quel mode (« parcourir index » ou « mots clé ») l'utilisateur utilise-t-il habituellement ?
- Question 3 : Quelle est la proportion des titres de monographies comportant un mot traité erronément comme un article initial par l'algorithme de repérage qui seront repérés par les usagers ? Cette proportion est-elle la même que pour les titres non affectés ?
- Question 4 : Les titres de monographies comportant un mot traité erronément comme un article initial par l'algorithme de repérage sont-ils en général plus ardu à repérer en termes de temps et d'effort que les autres titres ?

Pour répondre à ces quatre questions de recherche, nous avons effectué une analyse du comportement des utilisateurs lors de la consultation de l'index de titres d'un catalogue de bibliothèque. Dans un premier temps, nous avons obtenu de l'Université de Montréal, le journal des transactions de recherche effectuées en mode « parcourir index » dans l'index de titres pendant le mois d'octobre 2005 dans le catalogue Atrium. À partir de ces données, il nous a été possible de vérifier si en général l'utilisateur conserve ou omet les articles initiaux de ses requêtes lors de ses recherches par titre.

Pour répondre aux autres questions de recherche, nous avons d'abord recensé dans le catalogue de l'Université de Toronto tous les titres débutant avec un mot pouvant être considéré erronément comme un article initial. Nous avons ainsi constitué un fichier contenant le titre des documents susceptibles d'être plus difficilement repérables, c'est-à-dire les documents dont les titres débutent avec le mot «*a*», «*an*» ou bien «*the*» alors que ce mot n'est pas un article (p. ex. *À bout portant* ; *An der Wegscheide* ; *Thé ou café, Monsieur le Ministre* ?). Nous avons ainsi répertorié 4 384 titres de documents présentant ce type de problème. Il faut toutefois noter que pour constituer notre échantillon, seuls les titres en français et en anglais (c'est-à-dire 1 545) ont été retenus.

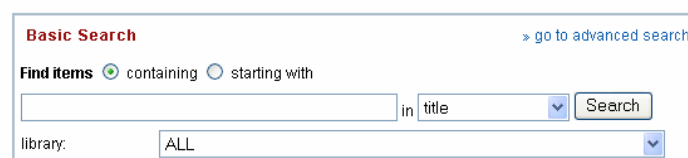
À partir de cet ensemble de titres problématiques, 24 listes de titres, chacune constituée de 30 documents sélectionnés de façon aléatoire, ont ensuite été préparées. Afin de ne pas influencer le comportement de recherche des participants lors de leur recherche, nous avons entremêlé différents types de titres à l'intérieur de chaque liste. Chaque liste de 30 titres était donc constituée de trois groupes de titres de cette façon :

- 1<sup>er</sup> groupe : 5 titres débutant avec un mot « ordinaire » (i.e., ni un article, ni un homographe d'article)  
p. ex. *Out after dark*
- 2<sup>e</sup> groupe : 10 titres débutant avec de véritables articles initiaux  
p. ex. *A very profitable war*
- 3<sup>e</sup> groupe : 15 titres « problématiques », c'est-à-dire débutant avec un mot homographe à un article initial  
p. ex. *À la plage.*

Tous les titres inclus dans les listes étaient présents dans le catalogue de l'Université de Toronto, donc en principe repérables. L'ordre de présentation des titres sur la liste des titres à rechercher a été déterminé de façon aléatoire et modifié pour chaque liste afin de minimiser le facteur d'apprentissage. Un exemple est présenté en annexe.

Une fois ces listes constituées, nous avons demandé à 24 étudiants de niveau pré-universitaire (1<sup>re</sup> ou 2<sup>e</sup> année de Cégep du profil pré-universitaire) de retracer les notices bibliographiques des titres énumérés sur une des listes, chaque participant recevant une liste différente pour éviter les effets de contamination. La collecte de données s'est déroulée au Collège de Maisonneuve (Montréal, Canada), du 30 janvier au 6 février 2006. Le recrutement s'est effectué à l'aide d'affiches expliquant les tâches demandées, le temps requis estimé, et la rémunération offerte. La principale raison pour sélectionner des étudiants de niveau collégial était d'assurer une certaine uniformité quant au niveau d'expérience et d'exposition à la recherche bibliographique dans les catalogues.

Lors de l'expérience, nous avons demandé à chaque étudiant de rechercher tous les titres se trouvant sur sa liste en utilisant, à son choix, l'un ou l'autre des modes de recherche *containing* ou *starting with* tel que le montre la figure 2 :



The image shows a web-based search interface titled "Basic Search". At the top right, there is a link that says "> go to advanced search". Below the title, there are two radio buttons: "containing" (which is selected) and "starting with". Underneath these is a search input field, followed by the word "in" and a dropdown menu currently showing "title". To the right of the dropdown is a "Search" button. At the bottom left, there is a label "library:" followed by a dropdown menu currently showing "ALL".

**Figure 2 — Interface de recherche de base du catalogue de l'Université de Toronto**

Mentionnons que nous avons alterné le mode de recherche en choisissant initialement le mode *containing* pour la moitié des participants et le mode *starting with* pour l'autre moitié, afin qu'aucun des deux modes ne soit privilégié, au moins au début du processus de recherche des participants. L'index *title* était sélectionné au préalable par défaut et ne devait pas être changé par les participants.

Chaque session de recherche a été enregistrée à l'aide du logiciel Camtasia, une application qui permet de consigner toutes les opérations effectuées à l'écran et de créer des vidéos reproduisant fidèlement les sessions de recherche. La figure 3 présente une image tirée d'une telle vidéo :

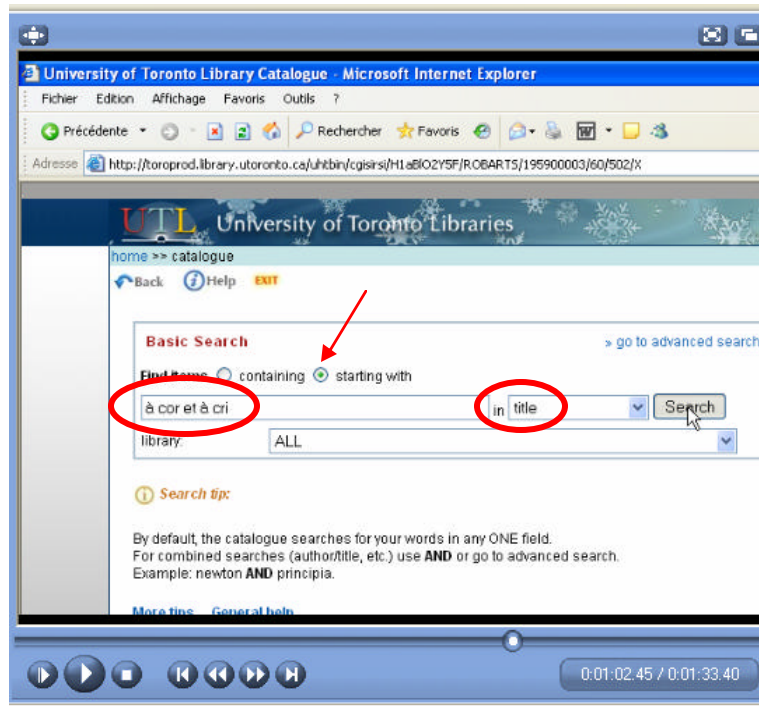


Figure 3 — Image extraite d'une vidéo créée avec le logiciel Camtasia

Une fois le document repéré les participants devaient inscrire la cote sur le formulaire prévu à cet effet (voir en annexe) ce qui nous a permis de calculer rapidement le taux de succès. De plus, pour chaque titre recherché, les variables suivantes ont été enregistrées :

- Temps de début : le moment où l'utilisateur exécute sa requête en cliquant sur le bouton *search*
- Temps de fin : le moment où l'utilisateur affiche la bonne notice (si trouvée)
- Mode de recherche pour chaque requête : mode *containing* ou *starting with*.
- Nombre de résultats : dans le cas des requêtes en mode *containing*, la taille de l'ensemble de résultats obtenus
- Inclusion ou omission de l'article initial pour les titres débutant par un article.

#### 4. Observations et analyse

Les données recueillies nous ont permis d'effectuer les calculs nécessaires pour répondre aux questions de recherche posées.

*Question 1 : Habituellement, l'utilisateur conserve-t-il ou omet-il les articles initiaux de ses requêtes lors de ses recherches en mode « parcourir index » dans l'index de titres ?*

L'analyse des requêtes recueillies dans le journal de transactions du catalogue Atrium révèle que les usagers semblent conserver les articles initiaux dans leur requête environ deux fois sur trois. Sur les 12 216 requêtes recueillies dans le journal de transactions, 1 468 requêtes (environ 12 %) étaient des requêtes effectuées pour chercher des titres ayant un article initial. Notre analyse s'est donc concentrée sur celles-ci. Nous observons que l'article initial n'a été omis que dans 36,8 % des cas, laissant croire que les usagers ont plus souvent l'habitude de conserver les articles initiaux dans leur requête. Kan et Poo (2005, 93) affirment également avoir observé ce phénomène.

Nombre de requêtes en mode « parcourir index titre » (total)	12 216	
Nombre de requêtes pour des titres ayant un article initial	1 468	(100 %)
Nombre de requêtes où l'article initial a été omis	540	(36,8 %)
Nombre de requêtes où l'article initial a été conservé	928	(63,2 %)

**Table 1 — Analyse du journal de transaction d'Atrium pour octobre 2005**

Des proportions similaires ont été observées lors de l'analyse des vidéos des séances de recherche dans le catalogue de l'université de Toronto. Ainsi, sur les 54 requêtes effectuées en mode « parcourir index » pour chercher des titres ayant un article initial, 37 (68,5 %) contenaient l'article initial tandis que dans les 17 autres (31,5 %), l'utilisateur avait omis d'inclure l'article. Toutefois l'analyse semble indiquer que les requêtes dans lesquelles les articles initiaux ont été omis sont plus souvent couronnées de succès. On constate en effet que des 37 requêtes dans lesquelles l'article a été conservé, seulement 18 (48,6 %) ont permis de repérer la notice recherchée. Ce taux grimpe à plus de 88 % lorsque l'article initial est omis de la requête.

Nombre de requêtes en mode « parcourir index titre » (total)	213		Requêtes
Nombre de requêtes pour des titres ayant un article initial	54	(100 %)	fructueuses
Nombre de requêtes où l'article initial a été omis	17	(31,5 %)	15 (88,2 %)
Nombre de requêtes où l'article initial a été conservé	37	(68,5 %)	18 (48,6 %)

**Table 2 — Analyse des requêtes des séances (catalogue de l'Université de Toronto)**

\* \* \*

*Question 2 : Pour effectuer des recherches de titres connus, quel mode (« parcourir index » ou « mots clé ») l'utilisateur utilise-t-il habituellement ?*

La compilation du nombre total de requêtes effectuées par les 24 participants pour chercher leurs 30 titres, indique que plus du trois quart des requêtes ont été effectuées en mode « mots-clés ». Cette proportion grimpe à plus de 80 % si on ne compile que la première requête effectuée pour chaque titre.

Mode « parcourir index »	234 (23,1 %)
Mode « mots-clés »	778 (76,9 %)
Total	1 012 (100 %)

**Table 3 — Nombre total des requêtes**

Mode « parcourir index »	128 (17,8 %)
Mode « mots-clés »	592 (82,2 %)
Total	720 (100 %)

**Table 4 — 1<sup>re</sup> requête effectuée pour chaque titre**

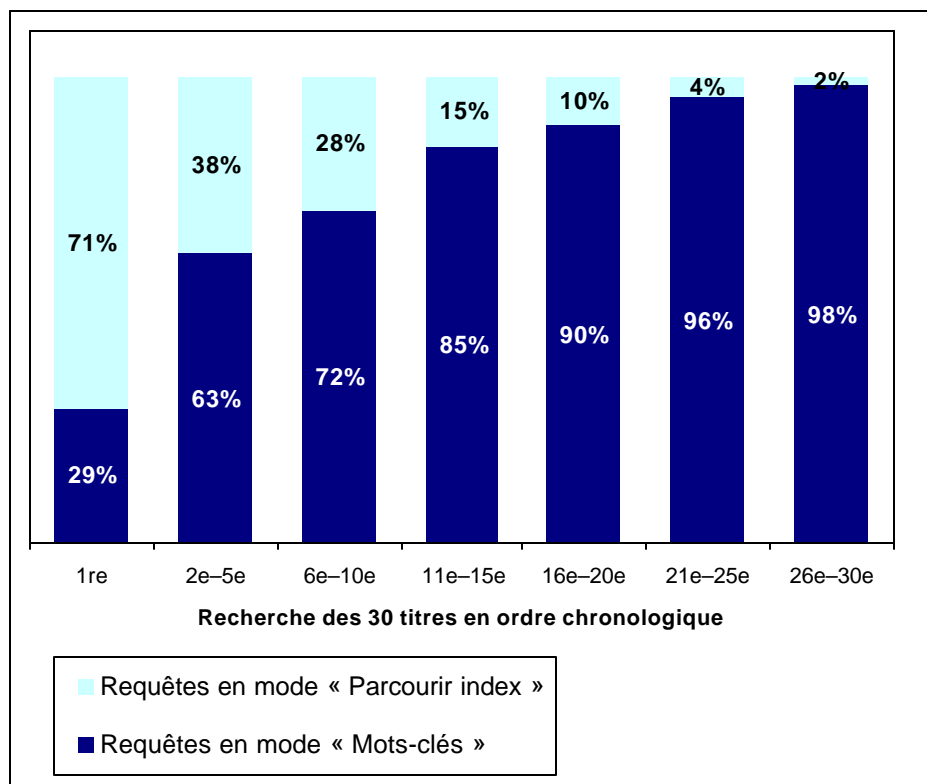
Suite à ces observations on pourrait conclure que le mode « mots-clés » est le mode qui est préféré par les usagers. Il faut toutefois se rappeler que les échantillons des titres soumis aux participants comportent 50 % de titres problématiques ce qui n'est pas du tout représentatif de la proportion réelle de ces titres dans un catalogue (moins de 0,4 % selon l'étude effectuée sur le catalogue Atrium (Arsenault & Ménard 2005, 7)). À cause de l'algorithme de détection automatique des articles initiaux, il est pratiquement impossible de repérer ces titres en mode « parcourir index ». L'analyse de nos données révèle en effet qu'aucune des 360 notices problématiques n'a pu être repêchée avec ce mode. L'analyse pour tous les titres révèle que neuf fois sur dix, la dernière requête (celle qui a

permis d'afficher la bonne notice) est une requête en mode « mots-clés ». Il n'est donc pas étonnant que les usagers aient en bout de ligne utilisé ce mode plus souvent.

Mode « parcourir index »	64 ( 9,6 %)
Mode « mots-clés »	600 (90,4 %)
Total	664 (100 %)

**Table 5 — Dernière requête effectuée pour chaque titre trouvé**

Si on fait une analyse chronologique des requêtes on se rend tout de même compte qu'au départ les usagers utilisaient plus le mode « parcourir index » ; en effet, 17 des 24 participants (71 %) ont utilisé ce mode pour effectuer leur toute première requête bien que nous avons pris soin, pour la moitié d'entre eux, de sélectionner initialement le mode « mots-clés » comme choix de départ.



**Figure 4 — Mode utilisé pour la première requête effectuée pour chaque titre**

En observant le graphique, on remarque qu'au début de la séance (pour les 5 premiers titres), les deux modes sont utilisés à une fréquence quasi similaire. Mais en cours de séance on remarque que les usagers délaissent progressivement le mode « parcourir index » au profit du mode « mots-clés » (seulement 2 % des requêtes en mode « parcourir index » pour les 5 derniers titres cherchés) bien que le mode « parcourir index » soit en principe plus efficace pour retracer un document connu. Il est donc difficile d'affirmer que les usagers préfèrent utiliser le mode « mots-clés » car ici la surreprésentation des titres problématiques fait en sorte de donner l'impression que le mode « parcourir index » est moins efficace.

\* \* \*

*Question 3: Quelle est la proportion des titres de monographies comportant un mot traité erronément comme un article initial par l’algorithme de repérage qui seront repérés par les usagers ? Cette proportion est-elle la même que pour les titres non affectés ?*

La recherche d’un document connu (*known-item search*) pour lequel le titre exact nous est fourni constitue sans contredit l’une des tâches les plus faciles que l’on peut imaginer dans un catalogue. On imagine que le taux de succès devrait être très près de 100 %. C’est effectivement ce qu’on observe pour les titres de notre échantillon qui ne sont pas problématiques. On remarque toutefois que pour les titres jugés problématiques en raison de la présence en début de champ d’un mot homographe à un article, en moyenne deux titres sur quinze ne sont pas retrouvés.

	Moyenne	Écart type
Titres non problématiques ( $n = 15$ )	14,7 (97,8 %)	0,56
Titres problématiques ( $n = 15$ )	13,0 (86,7 %)	1,69

**Table 6 — Titres retrouvés en moyenne par les 24 participants**

Une comparaison des moyennes obtenues à l’aide du test  $t$  révèle que les différences observées sont significatives ( $p < ,0005$ ). Nous avons donc observé que les titres jugés problématiques en raison de la présence d’un mot erronément traité comme un article initial par l’algorithme de repérage étaient plus difficiles à retrouver.

\* \* \*

*Question 4: Les titres de monographies comportant un mot traité erronément comme un article initial par l’algorithme de repérage sont-ils en général plus ardu à repérer en termes de temps et d’effort que les autres titres ?*

### Temps

Le temps qui a été mesuré est calculé entre le moment où l’usager lance sa requête (en appuyant sur le bouton « search » (excluant donc le temps de la saisie qui peut fluctuer en raison de la longueur variable des titres) et le moment où la notice désirée s’affiche à l’écran. Les titres qui n’ont pas été trouvés ont été exclus de la moyenne.

	Moyenne (secondes)	Écart type
Titres sans article ( $n = 5$ )	5,58	6,28
Titres avec article initial ( $n = 10$ )	9,32	5,99
Titres problématiques ( $n = 15$ )	19,76	10,14

**Table 7 — Temps moyen (en secondes) pour trouver un titre**

L’analyse du temps nécessaire pour trouver les notices révèle qu’il a fallu beaucoup plus de temps en moyenne pour trouver les titres problématiques. Il a fallu un peu plus de temps pour trouver les titres ayant un article initial que les titres sans article, mais l’analyse statistique révèle que cette différence est non significative ( $p = ,062$ ). Par contre l’analyse statistique des titres problématiques avec les deux autres groupes de titres (avec et sans article) démontre que les différences observées sont dans ce cas-ci significatives ( $p < ,0005$ ).

## Effort

Dans le cadre de notre étude nous avons utilisé deux mesures, outre le temps, pour représenter l'effort fourni par les participants pour retracer un titre : les requêtes utilisées en moyenne, et la taille moyenne des ensembles repêchés (pour les requêtes en mode « mots-clés »).

Il a fallu en moyenne un peu plus de requêtes pour trouver les titres ayant un article initial que pour trouver les titres sans article mais l'analyse statistique révèle que cette différence est non significative ( $p = ,489$ ). Par contre, l'analyse statistique comparant les titres problématiques avec les titres des deux autres groupes (avec et sans article) démontre que les différences observées sont significatives ( $p < ,0005$ ).

	Moyenne	Écart type
Titres sans article ( $n = 5$ )	1,18	0,37
Titres avec article initial ( $n = 10$ )	1,25	0,28
Titres problématiques ( $n = 15$ )	1,66	0,33

**Table 8 — Nombre moyen de requêtes par titre**

Également les usagers ont obtenu en moyenne des ensembles légèrement plus gros (offrant moins de précision) pour trouver les titres ayant un article initial que les titres sans article mais l'analyse statistique révèle que cette différence est non significative ( $p = ,763$ ). Par contre l'analyse statistique des titres problématiques avec les deux autres groupes de titres (avec et sans article) démontre que les différences observées sont dans ce cas-ci significatives ( $p < ,005$  et  $p < ,011$  respectivement).

	Moyenne	Écart type
Titres sans article ( $n = 5$ )	3,11	4,56
Titres avec article initial ( $n = 10$ )	3,31	3,08
Titres problématiques ( $n = 15$ )	54,85	77,57

**Table 9 — Taille moyenne des ensembles par titre**

Selon les deux mesures que nous utilisons, il appert qu'en moyenne il a fallu plus de temps et il a été plus ardu de retracer un titre problématique.

## **5. Conclusions**

Notre analyse a tout d'abord révélé que lors des recherches en mode parcourir pour les titres débutant par un article, seulement un tiers des requêtes ne contenaient pas l'article. En effet, nous avons observé qu'environ deux fois sur trois, les usagers conservent l'article initial dans leur requête bien que ceux-ci soient ignorés lors de l'indexation. À l'époque des catalogues imprimés (p. ex. sur fiche), il était impératif de faire abstraction des articles initiaux pour localiser un titre à l'endroit approprié. Dans un environnement en ligne ce problème peut être contourné en ajoutant une liste d'exclusion afin de retrancher automatiquement les articles initiaux des requêtes, même si cela, comme nous l'avons vu, peut entraîner d'autres types de problèmes de repérage. Dans un environnement électronique, les requêtes sont probablement souvent formulées par des actions de copier et coller, ce qui peut expliquer, en partie, la présence des articles initiaux dans les requêtes. Suite à nos observations nous constatons que les usagers n'ont plus le réflexe d'éliminer les articles initiaux. L'utilisation d'algorithmes de détection automatique de ces articles semble donc utile. Nous regrettons toutefois que ces algorithmes ne soient pas plus sophistiqués, surtout dans les environnements

multilingues. Nous déplorons également que les usagers ne soient pas mieux formés à la recherche documentaire. Avant l'arrivée des catalogues informatisés les usagers savaient qu'il était nécessaire d'éliminer les articles initiaux pour rechercher un titre. Pourquoi en serait-il différent aujourd'hui ? Une solution alternative pour régler ce problème pourrait être l'assouplissement des règles de classement. Un titre contenant un article initial pourrait par exemple être classé à la fois à l'article et au premier mot significatif. Cette solution est d'ailleurs déjà appliquée dans certains catalogues. Ainsi, il deviendrait inutile d'utiliser les algorithmes de détection automatique des articles initiaux car il serait possible de retrouver ces titres d'une façon comme de l'autre, avec ou sans l'article dans la requête.

Par la suite, notre analyse a révélé que les algorithmes de détection automatique des articles initiaux, bien qu'utiles dans certains cas, sont un obstacle pour le repérage des titres qui débutent avec un mot «normal» qui est homographe à un article de la liste d'exclusion. Nous avons effectivement observé que le taux de succès pour retrouver ces titres est significativement inférieur à celui pour retracer les autres titres, car les titres «problématiques» ne peuvent être retrouvés en mode «parcourir index». Le mode «mots-clés», est un bon substitut dans bien des cas mais il arrive que pour les titres courts ou contenant que des mots ayant une occurrence élevée dans le catalogue, le repérage soit à toute fin pratique impossible. L'analyse a également révélé qu'il était plus ardu de retrouver les titres problématiques en terme de temps mais aussi en terme d'effort. Ceci se traduit en un plus grand nombre de requêtes nécessaires en moyenne pour retrouver ces titres et également en une baisse de la précision pour la recherche en mode «mots clé», les ensembles formés étant en moyenne plus grands. Cette analyse confirme la nécessité de fournir les deux modes de recherche «parcourir index» et «mots-clés» (Frost et al. 2000). Lorsque l'un n'est pas fonctionnel les taux de succès et l'efficacité des recherches sont affectés. L'utilisation des algorithmes de détection des articles initiaux peut avoir une certaine utilité si les usagers conservent les articles dans leurs requêtes mais ils hypothèquent la recherche en mode «parcourir index» pour certains titres ce qui semble avoir des répercussions négatives sur le repérage de ces titres. Il nous semble donc souhaitable de développer des méthodes alternatives pour pallier ce problème.

## Références

- Arsenault, C. et É. Ménard. 2005. Le traitement des articles initiaux dans les catalogues de bibliothèques : une étude de cas. *Actes du 33<sup>e</sup> congrès annuel de l'Association canadienne des sciences de l'information, CAIS/ACSI*. [http://www.cais-acsi.ca/proceedings/2005/arsenault\\_2005.pdf](http://www.cais-acsi.ca/proceedings/2005/arsenault_2005.pdf)
- Bourne, C.P. 1975. Initial article filing in computer-based book catalogs: techniques, problems, and article frequency. *Journal of Library Automation*, 8 (3), 221–47.
- Frost, C.O., B. Taylor, A. Noakes, S. Markel, D. Torres et K.M. Drabenstott. 2000. Browse and search patterns in a digital image database. *Information Retrieval*, 1 (4), 287–313.
- Kan M.-Y. et D.C.C. Poo. 2005. Detecting and supporting known item queries in online public access catalogs. In M. Marilino, T. Sumner, F.M. Shipman III (éds.) *International Conference on Digital Libraries Archive. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*. Denver, Colo. : ACM, 91–99.
- Larson, R. 1991. The decline of subject searching: long-term trends and pattern of index use in an online catalog. *Journal of the American Society for Information Science*, 42 (3), 197–215.
- Nielsen, R. et J. Pyle. 1995. Lost articles: filing problems with initial articles in databases. *Library Resources and Technical Services*, 39 (3), 221–22.
- Seeman, C. 2002. RE: Skipping initial articles.  
Disponible à : <<http://innovativeusers.org/list/archives/2002/msg02463.html>>.

## Annexe : Exemple d'une liste des titres à rechercher remise aux participants

P-01

Nom : \_\_\_\_\_

Adresse : \_\_\_\_\_

Tél. : \_\_\_\_\_ Courriel : \_\_\_\_\_

Âge : \_\_\_\_\_ Sexe :  M  F Langue maternelle : \_\_\_\_\_

Département : \_\_\_\_\_ Cycle : \_\_\_\_\_ Année : \_\_\_\_\_

### Liste des titres à rechercher

Titre à chercher		Cote
1.	A mi son	
2.	Blessings of the table : mealtime prayers throughout the year	
3.	La mer au large : roman	
4.	An apple a day : a holistic health primer	
5.	The night after Christmas	
6.	À propos d'indépendance	
7.	A cor et a cri, pour cor in fa et piano	
8.	A propos of Lady Chatterley's lover, and other essays	
9.	Le beau baiser : roman	
10.	A l'aube	
11.	À la recherche de l'or perdu : de la thésaurisation à la fonction monétaire	
12.	Les femmes antillaises	

Titre à chercher		Cote
13.	Always a loser	
14.	L'An 1 des droits de l'homme	
15.	Jamais contente	
16.	Sur le chemin Craig	
17.	Les plus beaux de nos jours	
18.	A very profitable war	
19.	Un jour, je te tuerai : roman	
20.	The most agreeable vice	
21.	A la rencontre de Jacques Prévert : Fondation Maeght 4 juillet - 4 octobre 1987	
22.	A Clichy : épisode de la vie d'artiste : opera-comique en un acte	
23.	A la souche obscure des rêves : la dialectique de l'écriture chez Yves Bonnefoy	
24.	L'An mil	
25.	A sa seule gloire : fragments d'une autre vie : roman	
26.	A travers l'épopée canadienne	
27.	A comme amour	
28.	A Londres, à Lourdes et à Rome en l'an de grâce 1908 : lettres au clergé	
29.	Out after dark	
30.	Une mort très douce : récit	