

УДК 81.27

LE RÔLE DES COMPÉTENCES LIÉES AU BILINGUISME DANS LA REFORMULATION DES QUESTIONS DANS LA RECHERCHE D'INFORMATIONS SCIENTIFIQUES

Omar Larouk**Dr, Maître de Conférences des Universités Hors-Classe (MCFU-HC) de l'École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB)****Université de Lyon**

ENSSIB, 17 boulevard de 11 Novembre 1918. Villeurbanne-Lyon, 69623, France. larouk@enssib.fr

Marina V. Garanovich**Candidate ès Lettres, Maître assistante du Département de la Linguistique théorique et pratique****Université nationale d'Etat de la Recherche scientifique de Perm**

614990, Russie, Perm. 15, Boukirev. garamarina@yandex.ru

Aujourd'hui l'accès à la production scientifique devient un important facteur dans le processus de diffusion des connaissances. Le prix très élevé des éditions scientifiques établi par les éditeurs privés (Elsevier, Springer, Wiley etc.), oblige les utilisateurs académiques à exploiter le potentiel de l'accès ouvert aux publications scientifiques. Dans l'article on donne l'estimation de l'accès à l'information scientifique pour les utilisateurs académiques sur Internet et de l'utilisation réelle des ressources commerciales électroniques. En outre on analyse les méthodes de la recherche d'information, qu'utilisent les professeurs et les étudiants de Master. On présente des diverses compétences (techniques, bilingues, etc.), qui contribuent à la recherche relevante de l'information multilingue parmi les utilisateurs académiques. L'étude montre que les compétences bilingues de la recherche sont utiles et nécessaires pour l'optimisation de l'accès à l'information scientifique.

Les mots clés: Internet; la demande de recherche; l'information multilingue scientifique; bilinguisme; les compétences bilingues.

1. Introduction

Le monde académique de la recherche scientifique (enseignants, chercheurs, professionnels des bibliothèques, doctorants, étudiants, etc.) est en train de vivre une profonde transformation avec le développement des «humanités numériques» ou «**digital humanities**», et notamment le mouvement pour le «Libre Accès à l'information scientifique et technique». Cette impulsion, née au début des années 90, connaît une forte croissance avec la possibilité aux utilisateurs d'y accéder à une production scientifique variée et multidisciplinaire.

La publication d'articles dans les revues électroniques a pris une place grandissante dans le processus de diffusion des connaissances et de communication entre universitaires, ainsi que dans l'accroissement de la notoriété des auteurs, le classement des universités et la réputation des nations.

L'accès aux documents stockés sur les catalogues de bibliothèques ou sur des bases de données bibliographiques a permis une ouverture de l'information scientifique aux différents utilisateurs. Ce bouleversement a plusieurs origines qui se complètent :

1) émergence de nouvelles technologies numériques: web, diffusion en ligne, mondialisation, rapidité, facilité d'usage, réseaux sociaux, etc;

2) variété de documents scientifiques contenant des informations pertinentes, etc.

Dans cet article, nous allons vous présenter deux enquêtes pour mettre en évidence le rôle que jouent les **compétences bilingues** dans le cadre de la **recherche d'information multilingue** chez les chercheurs de l'université (enseignants, chercheurs, doctorants) et chez les étudiants de Master (BAC+5). Dans chaque population, nous allons voir les comportements d'accès à l'information

scientifique et notamment la recherche des documents bilingues (français/anglais) à partir d'une question tapée en français ou en anglais.

On se posera aussi la question sur les nouveaux défis dans la reformulation des requêtes: les utilisateurs ont-ils besoins **des compétences bilingues et même multilingues** pour optimiser une recherche d'information scientifique sur les catalogues de bibliothèques, sur les bases scientifiques, sur le web ou bien NON?

2. La crise de l'édition scientifique: problème de financement de la documentation universitaire

Nous avons étudié la situation à l'Université Claude Bernard LYON 1, suite à une demande du département de formation à la recherche bibliographique (URFIST). Le but est de voir comment les taux d'utilisations des contenus achetés chez les éditeurs (Elsevier, Springer, Wiley, Blackwell-Emerald, Kluwer, etc.) sont utilisés par les chercheurs? En effet, le coût d'abonnement dépasse **le million d'euros** par année pour le budget d'université, et voir aussi si les chercheurs consultent «réellement» les contenus numériques payants ou préfèrent les documents gratuits.

Face à cette augmentation des prix des revues scientifiques, beaucoup d'appels [Harnad 2006] ont été lancé pour donner naissance à un mouvement BOAI¹ (Budapest sur l'Open Access Initiative) en 2000, qui favorise le libre accès aux productions des chercheurs. Le but est de rendre accessible le maximum de documents produits par la communauté scientifique **sans aucune compensation financière**.

3. Analyse sur les pratiques informationnelles des Enseignants/Chercheurs/Doctorants de l'université de Lyon1

L'Université Claude Bernard Lyon 1 est une université multidisciplinaire scientifique et médicale comprenant 102 unités de recherche, employant environ 3500 personnes dont 2100 enseignants/chercheurs et 1300 doctorants (25 % étrangers). Elle accueille depuis 2006 plus de 30000 étudiants.

Depuis les années 90, la bibliothèque universitaire propose des accès à tous, avec un abonnement à 7516 revues numériques (journaux électroniques), à 1513 revues-papier, à une quinzaine de bases de données (bibliographiques, encyclopédies, articles complets en PDF). Cela représente un budget de plus d'un million d'euros par année pour la bibliothèque.

Pour connaître l'usage réel des ressources électroniques commerciales, nous avons donc lancé une enquête par messagerie, sur les pratiques de

recherche d'information, auprès des enseignants/chercheurs et des doctorants.

3.1. Caractéristiques de la population interrogée

Les personnes interrogées, une population de 2000 enseignants et chercheurs et 490 doctorants, ont reçu un mail contenant le lien pointant vers un questionnaire² sous forme électronique, contenant 54 questions. Nous avons reçu 390 réponses, c'est-à-dire un échantillon de 15,6 %. Ainsi 52 % des personnes interrogées étaient en sciences (Chimie, Mathématiques, Biologie, etc.) et 48 % en santé (Médecine, Pharmacie, etc.). Et 53 % d'entre eux étaient des enseignants-chercheurs (Professeurs, Maître-conférences, Médecins, Pharmaciens, Psychiatres, etc.), 12 % chercheurs (ingénieurs de recherche) et 35 % d'entre eux doctorants (préparation de thèse).

3.2. Comportement des chercheurs pendant les étapes de la recherche d'information scientifique

Alors 26 % étaient au début de leurs recherches, 47% au milieu de leurs recherches, et 16 % à la fin de leurs travaux de recherche. C'est-à-dire, un chercheur a besoin en permanence de la fraîcheur des informations et des publications actualisées pour contrer la concurrence. Pour ce travail de recherche documentaire: 69,5 % des personnes interrogées utilisent les bases de données, 66,5 % d'entre eux consultent des revues-papier, 17 % utilisent les archives ouvertes (portails web gratuits).

3.3. Enquête sur l'accès bilingue à la documentation numérique des universitaires

Les Bases de données (BDD) sont majoritairement en **anglais**, compte tenu de l'importance de la documentation numérique dans les domaines scientifiques. Les chercheurs doivent reformuler leurs questions en langue anglaise et ceci est confirmé par l'enquête:

- **L'anglais** est la langue la plus utilisée lors d'une interrogation de toutes les bases (payantes et gratuites), et représente à lui seul 53 %.

- **Français et Anglais**, les deux langues utilisées dans le même temps, représentent 46,5 %.

- **Français** représente à lui seul, seulement 16 %.

Autre exemple particulier, les enseignants et chercheurs ont recours au bilinguisme (franco-anglais) très souvent lors de la recherche de documents bilingues, ce sont surtout en :

- **Anglais** pour les articles de revues électroniques dans (96 %),

- **Français** et Anglais: des cours dans (40,5 %), des thèses et rapports dans (34 %), des présentations orales dans (21 %).

3.4. L'accès aux moteurs de recherche spécialisés gratuits

Depuis 2005, il y a eu plus de 860 000 téléchargements à partir des collections de revues et plus de 110000 interrogations dans les bases de données fournies par les éditeurs. On constate donc une utilisation intense des ressources électroniques à Lyon1. Le basculement des documents «papier» vers les documents «numériques» a commencé avec l'arrivée de la base «ScienceDirect» qui couvre actuellement plus de 50% des recherches.

A cause de ce coût financier très important de «ScienceDirect», les universités ont décidé de réduire le budget, en poussant les universitaires à déposer leurs articles et communications sur un portail web national financé par le ministère de l'enseignement supérieur, appelé **HAL (Hyper-Articles-en-Ligne)**³ en plusieurs langues (français, anglais, allemand, etc.). HAL contient plus de 280 552 documents en texte intégral (thèses, articles, etc.) publiés par les chercheurs francophones. Cette population «scientifique» utilise aussi les outils gratuits suivants comme:

- Google Scholar⁴ est utilisé à 54,5% par les enseignants et chercheurs de Lyon1.
- Les universitaires en santé utilisent deux bases (PubMed pour les notices bibliographiques et PubMed Central, une base d'archive gratuite).

4. Enquête sur l'accès bilingue à la documentation numérique des étudiants de Master (BAC+5)

La terminologie documentaire (expression syntaxique des questions) dans un système de recherche d'information (bibliothèques numériques, moteurs de recherche, bases de données spécialisées, etc.) est un élément linguistique essentiel pour récupérer les documents, et les informations pertinentes [Fidel et al. 1999; Larouk 1998; Livonen, Sonnenwald 1998; Vakkari 2000].

On se posera la question, dans cette seconde enquête, sur les nouveaux défis dans la reformulation des requêtes: les utilisateurs ont-ils besoins **des compétences bilingues et/ou multilingues** pour optimiser une recherche d'information scientifique ou NON?

Notre protocole est constitué d'une enquête (papier) et d'une évaluation quantitative des pratiques sur un ordinateur. Voici les facteurs qui nous souhaitons observer pendant les reformulations des questions:

a) **Compétences de formulation terminologique:** Récupération des mots-clés précis à partir de la question posée.

b) **Compétences en informatique de base:** Windows et navigation-web.

c) **Compétences en documentation-bibliographique:** Connaissances des sources

documentaires et les outils d'accès (moteurs de recherche, base documentaire, catalogue de bibliothèques) avant le lancement de la requête.

d) **Compétences en reformulation logique:** Reformulation terminologique des questions composées de mots-clés et des opérateurs booléens (AND, OR, ANDNOT).

e) **Compétences bilingues pour la reformulation:** Capacités multilingues pour traduire les questions du français vers l'anglais ou vers d'autres langues (Allemand, Russe, Arabe, Espagnol, etc.).

4.1. Caractéristiques de la population des étudiants et les résultats

Nous nous sommes intéressés aux étudiants ayant une formation supérieure (4 années à l'université) et se préparent à une cinquième pour obtenir un master. Notre panel était composé de 50 participants qui ont choisi mes enseignements dont la répartition est la suivante: (50 = 10 en Informatique + 18 en Science de l'Information + 22 en Langues étrangères).

- 20 % des étudiants interrogés étaient en Master «Système Informatique» (université Lyon-1);
- 36 % des étudiants étaient du Master «Science de l'Information et des Bibliothèques» (ENSSIB);
- 44 % d'entre eux étaient étudiants du Master «Langues Etrangères Appliquées» (université Lyon2).

Dans cette partie, nous analyserons les éléments affectant la Recherche d'Information Multilingue véhiculée par les questions. Les données ont été recueillies au moyen d'un «questionnaire-papier» distribué aux étudiants de Master. Chacun a répondu aux questions (Q₁, Q₂) suivantes:

Q₁: Trouver toutes les informations qui traitent de la grippe aviaire H1N1 en français et en anglais?

Q₂: Trouver tous les documents qui traitent des moteurs hybrides de TOYOTA en français et en anglais?

Tous les répondants utilisent l'ordinateur depuis leur entrée à l'université, et presque tous les jours, dans les différents enseignements, ou à la maison. Ils maîtrisent toutes les fonctionnalités de base d'un ordinateur avec Windows, la messagerie, et la navigation sur le web, via le moteur Google.

4.2. Analyse empirique des réponses du questionnaire et tests sur les outils numériques

Tous les participants ont lu les instructions très simples du questionnaire avant de le remplir de façon indépendante (pas de communication entre eux!). Les questions ont été données en français, mais les réponses doivent contenir toutes les

formulations possibles et correctes dans toutes les langues (français, anglais, etc.).

Les 50 répondants ont rempli le questionnaire, et on a exigé au moins une réponse par question sur le papier et qu'ils écrivent les autres requêtes possibles, avant de les tester sur l'outil de recherche d'information.

4.2.1. Compétences en formulation terminologique.

Etape 1: Analyse terminologique des mots-clés issus des réponses du questionnaire.

Les résultats terminologiques observés sur les requêtes formulées par les 50 répondants lors de la première séance (15 minutes au maximum) sont:

- 3/10 (30 % en informatique) ont trouvé la bonne requête formelle avec les mots-clés;
- 4/22 (18 % en Langues Etrangères Appliquées) ont trouvé la bonne requête formelle avec les mots-clés;
- 12/18 (66 % en Science de l'Information et des Bibliothèques) ont trouvé la bonne requête formelle avec les mots-clés.

A l'issue de ces résultats, le but était d'harmoniser les réponses pour tous les étudiants afin qu'ils puissent entamer une recherche d'information sur ordinateur, avec une formulation terminologique **correcte** contenant les bons mots-clés.

Etape 2: Tests de recherche d'information via des outils numériques (catalogues, bases ou web).

Sur la base des résultats de l'enquête par questionnaire, et après les corrections pour tous, les étudiants ont réalisé des recherches d'informations sur: (le web, les BDDs et les catalogues de bibliothèques). On a laissé l'initiative aux étudiants de choisir l'outil approprié (catalogues ou base ou web) selon leurs connaissances.

Le but visé est que chacune des formulations permettait de trouver au moins une réponse pertinente à la question posée. Les étudiants ont fait la recherche d'information à l'aide d'un ordinateur, pendant 15 minutes, et voici les facteurs qui nous avons observé pendant les reformulations des questions.

4.2.2. Compétences informatiques de base: Windows et navigation-web. Tous les étudiants maîtrisent les fonctionnalités de Windows et du Web (Google, etc.), mais les meilleurs experts sont les informaticiens qui exploitent les fonctionnalités poussées du web et du système d'exploitation (Windows). C'est ce que, on appellerait **l'effet d'endurance**, d'un même outil manipulé depuis des années.

4.2.3. Compétences documentaires bibliographiques: Connaissances des sources et des outils d'accès (moteurs de recherche, base

documentaire, catalogue de bibliothèques) avant le lancement de la requête.

La connaissance des sources (fonds documentaires) et le type de documents recherchés sont en liaison avec le sujet de la recherche. Les étudiants de l'ENSSIB (Science de l'information) ont pu choisir les bons outils bibliographiques, suite à leur formation. Les autres étudiants n'ont pas une connaissance approfondie des fonds documentaires et les outils dédiés existants.

4.2.4. Compétences logiques de formulation: Reformulation terminologique des questions avec des mots-clés et des opérateurs booléens (AND, OR, ANDNOT)

Seuls les étudiants de l'ENSSIB ont exploité des opérateurs de recherche logiques comme (AND, OR, NOT, NEAR, etc.), pour avoir des requêtes précises et filtrées. Ces opérateurs n'ont pas été utilisés par les étudiants en langues, mais, deux étudiants en Informatique ont surpris leurs collègues en utilisant un opérateur logique (AND).

4.2.5. Compétences bilingues pour la reformulation: Compétences multilingues pour traduire les questions du français vers l'anglais, ou vers d'autres langues (Allemand, Russe, Arabe, Espagnol, etc.)

Seuls les étudiants en langues (LEA) ont proposé des traductions correctes des mots-clés en anglais, sans aide des outils de traductions en ligne. Les reformulations étaient pertinentes pour trouver le maximum de documents et d'informations correctes. Grâce à leurs compétences bilingues, ils ont trouvé les termes «sémantiquement» proches aux mots-clés. Cette expertise a contribué dans la formulation à produire des équivalences, et à utiliser ces termes synonymes.

Les autres étudiants ont cherché à traduire les requêtes, à l'aide d'outils en ligne, comme les dictionnaires sans arriver à obtenir les réponses pertinentes. Certains étudiants en science de l'information ont fini par trouver quelques réponses pertinentes, après un laps temps.

5. Discussion sur les résultats

Nous pensons que notre étude n'a pas révélé tous les effets liés la reformulation multilingue de la requête, car les compétences ont été évaluées uniquement sur deux requêtes (2 questions posées). Il serait plus intéressant et plausible de lancer une opération d'évaluation sur plusieurs questions dans le futur, afin d'observer la stabilité ou non des facteurs liés aux différentes compétences, et notamment au bilinguisme.

En effet, les possibilités de traduction d'une question en français vers l'anglais nécessitent des compétences poussées en syntaxe et en sémantique,

mais aussi au contexte de l'indexation du document (au moment de la création qui est liée à la spécificité du thème ou du domaine scientifique). D'ailleurs, Belkin a bien illustré le défi des utilisateurs face à la recherche d'information basé sur le texte dans «Helping People Find What They Don't Know» [Belkin 2000], c'est-à-dire «Aider les utilisateurs à trouver ce qu'ils ne savent pas», mais le problème déterminant est lié à la représentation des textes linguistiques par les systèmes informatiques, d'où l'importance d'une analyse linguistique avant l'indexation terminologique des textes (monolingues ou bilingues) et un traitement informatique de la langue prenant en compte plusieurs contextes (scientifique, sociolinguistique, etc.).

6. Conclusion

On peut faire un bilan en disant que les chercheurs universitaires, qui favorisent l'accès direct au texte intégral de l'article complet (*full-text*), sont aussi:

- obligés d'utiliser le bilinguisme pour trouver les articles dans les revues électroniques, les actes de conférence, les bases «Brevets» ou tout «document actualisé»;

- poussés à exploiter les potentialités des portails gratuits et de l'open accès académique (HAL, GS,...) à cause des coûts exorbitants exigés par les éditeurs privés (Elsevier, Springer, Wiley,...).

- Cependant, pour les étudiants de master, nous pensons qu'une formation à la méthodologie d'accès aux techniques bibliographiques est obligatoire, car il y a:

- une méconnaissance assez générale des ressources disponibles (types de sources documentaires);

- une inexpérience des outils bibliographiques de recherche d'information;

- une ignorance des langages documentaires (logique de Boole ou opérateurs AND, OR, ANDNOT).

Suite à nos observations, nous recommandons qu'une formation au **bilinguisme** ou **la traduction scientifique** pour tous les utilisateurs (universitaires, doctorants, étudiants) est une nécessité. En effet, nous avons vu que le choix des mots-clés est important, et seuls **les étudiants bilingues** formulaient des requêtes exactes en anglais (inter-langues), sans dictionnaire de traduction.

Notes

¹La communauté de scientifique et de bibliothécaires a abouti à des déclarations en faveur du mouvement libre et gratuit, comme la déclaration du BOAI (Budapest sur l'Open Access Initiative).

Elle encourage les chercheurs et institutions universitaires à adopter les serveurs d'archivage et les revues en libre accès (gratuits) pour publier leurs articles ou leur production scientifique.

²L'enquête a été menée par trois chercheurs de l'Université de Lyon (J-P. Lardy, S. Dalhoumi, O. Larouk), via un questionnaire, sur une période de 6 mois.

³<http://hal.archives-ouvertes.fr/>, ce serveur HAL est localisé à Lyon (HAL-CSDD-Lyon).

⁴GOOGLE SCHOLAR: <http://scholar.google.com/> moteur est un moteur de recherche spécialisé dans la littérature universitaire. Il indexe des articles issus de périodiques (revues gratuites), des thèses, des livres, des rapports qui proviennent d'éditeurs scientifiques, ou de serveurs universitaires, etc. Les chercheurs consultent, aussi, le moteur de recherche SCIRUS développé par l'éditeur Elsevier: <http://www.scirus.com>.

⁵Il faut préciser que les étudiants en Science de l'Information et des Bibliothèques ont eu des cours sur la recherche d'information scientifique au cours de leurs cursus.

Bibliographie

Belkin N.J. Helping People Find What They Don't Know // Communications of the ACM. 2000. V. 43. № 8. P. 58–61.

Fidel R. et al. A visit to the information mall: Web searching behavior of high school students / Fidel R., Davies R.K., Douglass M.H., Holder J.K., Hopkins C.J., Kushner E.J., Miyagishima B.K., Toney C.D. // Journal of the American Society for Information Science (=JASIS). 1999. V. 50(1). P. 24–37.

Harnad S. Generic Rationale for University Open Access Self-Archiving Policy. 2006. [Ressource électronique]. URL: <http://eprints.ecs.soton.ac.uk/12078> (дата обращения: 10.04.2014).

Larouk O. Modeling users needs: Schema of interrogation and filtering the answers from the WEB in mode Co-operation // Advances in Knowledge Organisation. №6. Würzburg: Ergon. 1998. P. 105–115.

Livonen M., Sonnenwald D.H. From translation to navigation of different discourses: A model of search term selection during the pre-online stage of search process // Journal of the American Society for Information Science (=JASIS). 1998. V. 49(4). P. 312–326.

Vakkari P. Cognition and changes of search terms, and tactics during task performance: A longitudinal study // Proceedings of the RIAO'2000 Conference. P.: C.I.D., 2000. P. 894–907.

THE ROLE OF BILINGUAL SKILLS FOR FORMULATING RETRIEVAL REQUEST WHILE SEARCHING FOR SCIENTIFIC INFORMATION

Omar Larouk

Professor of High School of Informatics and Librarianship
Lion University

Marina V. Garanovich

Senior Lecturer of Theoretical and Applied Linguistics Department
Perm State University

At present the access to scientific literature becomes an important factor in the process of knowledge spreading. High price of scientific publications set by private publishers (Elsevier, Springer, Wiley etc.) makes academic users exploit the potential of free access to scientific publications. The paper estimates access to scientific information for academic users in the internet and factual usage of commercial electronic sources. Methods of information search used by teachers and graduate students are analyzed. Various skills (technical, bilingual, etc.) that aid relevant search of multilingual information among academic users are discussed. The research proves that skills of simultaneous bilingual search are useful and necessary for optimization of access to scientific information.

Key words: Internet; retrieval request; multilingual scientific information; bilingualism; bilingual skills.

РОЛЬ ДВУЯЗЫЧНЫХ НАВЫКОВ ДЛЯ ФОРМУЛИРОВАНИЯ ПОИСКОВОГО ЗАПРОСА ПРИ ПОИСКЕ НАУЧНОЙ ИНФОРМАЦИИ

Омар Ларук

д. техн. н., профессор Высшей школы информатики и библиотечного дела
Лионский университет

Марина Владимировна Гаранович

к. филол. н., старший преподаватель кафедры теоретического и прикладного языкознания
Пермский государственный национальный исследовательский университет

Сегодня доступ к научной литературе становится важным фактором в процессе распространения знаний. Высокая цена научных изданий, устанавливаемая частными издателями (Elsevier, Springer, Wiley и т.д.), заставляет академических пользователей эксплуатировать потенциал открытого доступа к научным публикациям. В статье дается оценка доступа к научной информации для академических пользователей в интернете и фактического использования коммерческих электронных ресурсов. Кроме того, анализируются методы информационного поиска, которые используют среди преподаватели и магистранты. Демонстрируются различные навыки (технические, двуязычные и др.), способствующие релевантному поиску многоязычной информации среди академических пользователей. Исследование показывает, что навыки поиска одновременно на двух языках полезны и необходимы для оптимизации доступа к научной информации.

Ключевые слова: Интернет; поисковый запрос; многоязычная научная информация; билингвизм; двуязычные навыки.