

## **21. La BnTA, une idée portée par l'évolution des usages**

*C'est en centaines de millions de mots-occurrences que nous pouvons capitaliser le contenu de la BnTA. Le chiffre des pistes et des modes d'exploration de cette base est aussi impressionnant de par les entrées et les combinaisons que chaque utilisateur adoptera. Notre présent article s'intéresse à cette simultanéité des possibles.*

### **Plan**

1. Préliminaires
2. Les étapes
3. Les impacts
4. Bibliographie

### **1. Préliminaires**

L'émergence des plates-formes de livres numériques n'aurait pu avoir lieu sans le développement d'usages nouveaux, spécifiques à la recherche sur la toile. L'essor rapide de l'internet a en effet entraîné des changements profonds dans les modes d'accès au savoir et à l'information. Deux types d'usages expliquent en partie l'intérêt suscité par le développement de bibliothèques numériques et peuvent profondément influencer les réflexions en matière d'élaboration de telles bibliothèques virtuelles et de bases de données. Le premier de ces usages est le recours désormais prioritaire aux moteurs de recherche. Le second relève de la masse toujours grandissante d'informations à la disposition des internautes.

Les moteurs de recherche sont aujourd'hui des outils universellement reconnus comme particulièrement efficaces pour permettre aux internautes d'accéder à la masse de connaissances disponibles sur la toile.

Deux éléments définissent l'efficacité dans une recherche au niveau d'une base de données : la pertinence et la puissance.

L'initiative de Google, de numériser les ouvrages sans avoir toujours les accords préalables avec les partenaires et d'en garder l'exclusivité a fait réagir les Etats au travers d'organisations institutionnelles. C'est une précaution particulièrement importante qu'il s'agira de mettre en avant en élaborant une charte.

Les modalités d'interrogation et de restitution des textes numérisés dépendent essentiellement des accords passés entre les partenaires.

La possibilité de mettre des livres intégralement interrogeables et pouvant être lus dans leur intégralité sera réservée par exemple à des institutions ayant procédé à des abonnements. Cependant, des extraits d'ouvrages, sous forme de Snippets, pourraient être consultés librement sans porter atteinte aux droits d'auteur. Signalons que l'évolutivité des fichiers est envisageable car il semble probable que les chaînes d'océrisation et de corrections, a posteriori des textes par traitement de masse, vont se perfectionner dans les années à venir.

La mission de la base de données outre qu'elle offre une base consultable joue aussi le rôle de conservation de la production nationale, même si les processus actuels de numérisation de masse restent à perfectionner. Une conservation pérenne

numérique répondra à bien des problèmes liés au stockage tout en tenant compte de l'évolutivité des formats informatiques.

## **2. Les étapes**

L'élaboration de la base de données telle qu'elle est prévue dans le présent projet constitue une partie de ce qui devrait être un vaste chantier de numérisation de masse, il faut toutefois insister sur le travail accompli car il s'agit de scanner chaque page de chaque exemplaire, sans destruction ni détérioration de l'ouvrage, celui-ci devant être restitué après traitement à l'institution qui le conserve pour réintégration dans les collections. Aucun ouvrage n'est massicoté : les pages sont tournées à la main par un opérateur (plusieurs milliers de pages par un individu)

Liste de propositions

Une mission sur la numérisation des fonds culturels patrimoniaux devrait se constituer et réunir les protagonistes des domaines concernés pour procéder à des analyses dans le cadre technique, économique et juridique dans lesquels s'inscrivent traditionnellement les modes de gestion et de fonctionnalité.

Une organisation en métadonnées : Les métadonnées sont des éléments de description des ouvrages numérisés.

## **3. Les impacts**

La constitution d'une base de données « Bnta », véritable source d'information qui contribue à la collecte des ressources documentaires pour l'Algérie en matière de texte.

- Sur le plan socio- économique, la Bnta devient un index de référence incontournable aux différents partenaires nationaux et internationaux.
- Sur le plan intellectuel, cette Bnta devient la Banque Numérique Nationale donnant accès à l'information, la préserve et la rend pérenne.
- Sur le plan universitaire, cette base de données devient un outil d'exploitation en termes de ressources bibliographiques, mais encore un espace reconnu par les universités algériennes et internationales.

Choix des corpus, combinaisons des commandes, chronologie des documents, listes des redondances, hapax, nullax, spécificités et bien d'autres constituent une illustration des applications de la BnTA. Le parcours de sens se trouve alors enrichi des observations, des clichés et tics d'auteurs, de leurs cryptages, de leur génome. Qu'il s'agisse d'un traitement thématique, lexical ou morphosyntaxique, l'observation et le traitement statistique ouvrent de nouvelles logiques d'interrogation en assistant, en guidant, en documentant le chercheur chevronné comme le simple utilisateur.

## **Bibliographie**

Béhar H., "Un projet de banque de données d'histoire littéraire" in Méthodes quantitatives et informatiques dans l'étude des textes, Genève, Slatkine, Paris : Champion 1986, p.43-54.

Bernard M., "La banque de données d'histoire littéraire", Bulletin de l'EPI, sept. 1988, N° 51, p.172-177.

Bernard M., "La banque de données d'histoire littéraires" in Les banques de données littéraires comparatistes et francophones, Limoges : Pulim, 1992, p.255.

Bernard M., Introduction aux études littéraires assistées par ordinateur, Paris, Puf, 1999. Charaudeau P. , Maingueneau D., Dictionnaire de l'analyse du discours, Seuil, 2002.

Lebart et Salem, Statistique textuelle, Dunod, Paris, 1994.

Müller C., Principes et méthodes de la statistique lexicale, Paris, Champion  
Collection Uni, 1992.

Müller C., Initiation aux méthodes de la statistique lexicale, Paris, Champion  
Collection Uni, 1992