

## **13. Constitution d'une banque de données textuelles, la BnTA**

*C'est une volonté marquée que de vouloir constituer une base de données des textes algériens. La technologie contemporaine permet d'arriver à cette immense collecte de notre patrimoine culturel oral et écrit et d'en conserver son extrême densité.*

### **Plan**

1. Définition
  2. Les types de bases, exemples de bases spécialisées
  3. Domaine d'exploitation
  4. « Un placement » pédagogique
  5. La feuille de route
  6. Frantext, Québetext, ou plus à côté Ibn Rushd
- Bibliographie

### **1. Définition**

Une banque de données est un ensemble de données relatif à un domaine défini de connaissances et organisé pour être offert aux consultations d'utilisateurs. Il existe plusieurs types de banques de données, des banques spécifiques ou plus générales comme c'est le cas de notre projet.

### **2. Les types de bases, exemples de bases spécialisées**

On distingue les banques de données bibliographiques (références de documents primaires, avec ou sans résumé), les banques de données iconographiques (images fixes ou animées ; à ne pas confondre avec le mode image), les banques de données textuelles (texte intégral complet ou partiel de documents primaires) ou document complet au format original, documents comprenant des textes, des images et des sons.

Exemple de bases de données spécialisées :

*L'Encyclopédie Médico-Chirurgicale (EMC)* est constituée d'une quarantaine de traités à fascicules mobiles qui couvrent l'ensemble des spécialités médicales. L'information est mise à jour trimestriellement. Chaque article étant rédigé par un spécialiste du sujet et validé par un comité de rédaction.

Les exemples de Frantext ou Québetext bases de données sont plus proches de nous dans la mesure où il s'agit de banques de données de bases textuelles. Nous y reviendrons en détail.

### **3. Les domaines d'exploitation**

Les domaines d'exploitation sont variés, notamment, un certain nombre de domaines fondamentaux comme le droit, le commerce, les affaires l'histoire, l'économie, la philosophie, le droit, les techniques, les religions allant jusqu'aux informations sur la vie quotidienne des populations, la littérature occupant pour nous une place privilégiée.

Une banque de données d'histoire littéraire trouve sa nécessité, et tire sa légitimité dans le fait de pouvoir répondre à une demande que seule l'informatique peut satisfaire permettant des fouilles approfondies. La récurrence ou non de thèmes, les variations des points de vue qui y sont portés, aussi bien dans l'espace que dans

le temps, par les personnages et leurs auteurs, les réflexions des critiques, les cadres culturels décrits, sont autant d'éléments dont l'analyse ne manquera pas de se révéler payante. Ainsi, des études comparatives, des analyses contrastives, lexicographiques peuvent être menées sur des corpus de très grande taille. Des mouvements esthétiques sont discernés d'une manière pertinente, la base de données fournissant les supports synchroniques et diachroniques. L'évolution de l'écriture d'un écrivain, d'une notion ou même de simples unités linguistiques significatives seront aisément repérables avec un support comme une banque de données. Des études stylistiques, terminologiques, des analyses génériques, des études thématiques, chronologiques, pour cerner le « génome » d'écrivains, pour extraire des idéologies sous-jacentes...

Si la base de données contient plusieurs sous divisions il est alors possible de constituer immédiatement un corpus d'œuvres qui traitent de thèmes et de sujets correspondants, d'interroger cette base et d'en tirer des analyses susceptibles d'apporter des solutions comme l'élaboration de manuels scolaires sur tous les niveaux d'enseignement, d'apporter des modifications de programmes et même de relever les points saillants d'une société comme modes de représentation collectifs et/ou imaginaires....

L'on sait que la juste résolution de problèmes aussi importants conditionne le développement d'un pays.

#### **4. « Un placement » pédagogique**

Outil d'investigation par sa fonction documentaire, une banque de données textuelles peut être aussi d'une grande utilité au plan pédagogique.

Regardons de près l'exemple de ma BDHL (Banque de Données de d'Histoire Littéraire), Michel Bernard explique à ce propos :

*« Tout d'abord nous familiarisons les étudiants avec le matériel et les procédures informatiques. D'autre part l'approche pédagogique s'est trouvée bouleversée. L'étudiant navigue librement dans l'information, et y cherche des renseignements qui ne lui sont plus communiqués, mais qu'il a dû constituer en savoir... Nous demandons à nos étudiants de nous aider à compléter la B.D.H.L. qui est bien sûr un système ouvert et évolutif. Là encore cette pédagogie active a donné les meilleurs résultats »*

Outre les supports papiers qui sont nécessaires, il y a aussi le fait que l'une des tendances pédagogiques actuelles, est l'enseignement assisté par ordinateur. Depuis l'année 1990, la B.D.H.L. française est utilisée à l'Université de la Sorbonne Nouvelle - Paris III comme support d'une unité d'enseignement obligatoire de première année d'initiation à l'histoire littéraire

Dans cette perspective notre projet d'une banque de textes algériens servira de support pour le renouvellement de l'enseignement de la production algérienne mais aussi des littératures africaines d'une façon générale, l'AUF ayant avancé l'idée d'une base de données de textes africains de langue française.

#### **5. La feuille de route**

Comment faire ?

Collecte de données/ Constitution de corpus

Numérisation

Correction des documents numérisés

Encodage

Certaines banques de données de textes sont déjà en cours de réalisation :

Impacts :

La constitution de ces banques permettra la mise en réseau des données, et leur diffusion sur l'autoroute de l'information, en interaction avec les banques de données

existantes, et celles qui seront créées dans d'autres espaces. Notre projet amènerait donc notre base de données à faire partie intégrante d'un réseau plurisectoriel de banques de données des textes francophones et/ou autres, destiné à mettre en interaction les données de ces banques par mode de croisements et de requêtes par tri multiple...

Les utilisateurs pourront interagir, s'interroger, se fournir les éléments de connaissances, de pratiques et d'informations de tous ordres, s'interpeller, se corriger, s'enrichir les uns les autres d'une façon presque instantanée. L'interactivité concernera donc les données elles-mêmes et leurs usagers ou utilisateurs.

Avec ces dispositions la recherche connaîtra certainement un développement sans précédent, en gagnant en précision et en efficacité.

Cette capacité d'analyse que permet une banque de données devient fructueuse quand on l'étend à plusieurs niveaux, par la mise en relation des données linguistiques, lexicales, contextuelles, thématiques, avec les données factuelles, historiques biographiques, politiques et sociales, dans une perspective comparatiste. L'intérêt des banques de données d'histoire littéraires est encore plus manifeste pour l'étude des genres littéraires qui relèvent des arts du spectacle, comme le théâtre et le conte.

Avec maintenant la possibilité de graver l'enregistrement filmé d'une œuvre dramatique ou la mise en scène d'un conte sur CD-ROM, *"pour la première fois dans l'histoire littéraire, souligne, Roger Odin, l'histoire du théâtre peut être envisagée sérieusement comme une histoire des "représentations" théâtrales : étude du style d'un metteur en scène, comparaison entre les différentes mises en scènes d'une même pièce, etc."* De la même façon *"le conte peut être appréhendé dans toutes ses dimensions : textuelle, orale, kinésique (étude de la gestualité), proxémique (étude de la situation spatiale) et communicationnelle"*

Les littératures écrites et orales africaines, trouveront dans ces nouveaux moyens de conservation et de traitement qu'offrent l'informatique et les banques de données d'histoires littéraires, des supports beaucoup plus fiables que les bandes magnétiques, et un environnement plus propice à leur épanouissement. La recherche scientifique qui, dans ses dimensions littéraire et culturelle, a sur tous ces plans un rôle moteur, aura d'autant moins de peines à s'acquitter de ces tâches qu'elle aura déjà gagné en célérité et efficacité.

L'élaboration de cet outil exige une attention soutenue, une masse considérable de travail, qui peut faire l'objet d'un programme de recherche décomposable en plusieurs opérations distinctes, quoique complémentaires.

La toute première tâche consistant à établir la liste des auteurs et des œuvres et des textes préconisés, et à faire un état précis des lieux qui doit être un inventaire exhaustif de tout ce qui a déjà été produit : ouvrages d'ensemble, monographies sur les auteurs et les œuvres, études d'un genre, d'une période, réflexions même partielles touchant d'une manière ou d'une autre aux textes concernés, certains impératifs s'imposent. Sur ce plan, la production littéraire algérienne n'échappe pas à cette réalité commune aux littératures africaines, qui est la faiblesse de la documentation critique. Il s'agira à ce stade, pour le chercheur, non seulement de la repérer dans l'espace, mais aussi et surtout de l'organiser en dressant une bibliographie. Ce premier document a toute son importance, parce qu'il conditionne la phase suivante qui est la collecte des données de base. Une des principales difficultés réside dans l'indisponibilité d'une bonne partie des œuvres dans les bibliothèques et les librairies. Cela résulte d'une part, de la grande place qu'occupe dans la publication, l'édition à compte d'auteurs qui est de diffusion restreinte, et de l'autre du non-respect de l'exigence du dépôt légal. Un grand nombre de textes algériens se trouve dans cette situation. La seule possibilité est de recourir aux

emprunts des textes à des fins de duplication. Il faut aussi réunir toute la documentation critique sous toutes ses formes : textuelle, iconographique et audiovisuelle. Cette opération pose les bases de ce qui deviendra une bibliothèque spécialisée, pour la littérature étudiée, et ouvre la voie à la rédaction des textes qui constitueront de la banque de données.

Le public, l'infrastructure éditoriale, les médias de diffusion, les canaux de circulation et de commercialisation du livre, ses conditions et espaces de réception sont autant de paramètres dont il faudra tenir compte. Il faut préciser que certaines opérations peuvent être le fait d'un seul chercheur, mais la plus grande partie fait l'objet d'un travail en équipe. Bien entendu une phase de formation et d'expertise en matière de constitution de banque de données textuelles seront établies : plusieurs points restent à développer.

## **6. Frantext , Québectext, ou plus à côté Ibn Rushd**

Quelques précautions indispensables :

L'expérience de Virginia Coulon est significative

*"la conception et la réalisation d'une banque de données s'appuient à tout moment sur deux idées clés : la prévision et l'analyse. Ceci est très important dès la genèse du projet...Prévoir large ! Etre ambitieux au départ. Prévoir toutes les évolutions possibles du projet, même si elles ne sont pas exploitées dans l'immédiat."*

Exemple avec Frantext

Frantext est le nom d'une base de données de textes français : textes littéraires et philosophiques, mais aussi scientifiques et techniques. Elle a été créée autour d'un noyau de mille textes, dans les années 70, afin de fournir des exemples pour le *Trésor de la Langue Française*. Une fois le dictionnaire terminé, elle a continué à évoluer : d'abord mise à disposition dans les années 80 sous forme d'un CD ("Discotext"), elle a été mise en ligne sur le web en 1998 par l'informaticien auteur de son moteur de recherches, Jacques Dendien.

Elle est régulièrement enrichie et comporte aujourd'hui près de 4000 références. Elle est la seule à proposer des recherches sur des textes qui vont de 1180 à 2009, avec un fonds contemporain particulièrement riche (820 textes sont postérieurs à 1950). Elle se décline en plusieurs bases :

- Frantext général (totalité des textes, accès par abonnement)
- Frantext catégorisé (1200 textes étiquetés grammaticalement, accès par abonnement)
- Frantext agrégation
  - Frantext Moyen Français
- Frantext Démonstration (une sélection de trente-cinq textes libres de droit, accès libre et gratuit).

S'agissant par exemple de l'analyse textuelle, Etienne Brunet, pour donner une idée des possibilités qu'offre Frantext, note qu'une fois la liaison télématique établie avec le serveur de Nancy,

*"le chercheur définit son corpus, en précisant à sa guise, les dates, le genre, le ou les auteur(s), le ou les texte(s) désiré(s), ou toute combinaison de ces critères. Puis il choisit les mots ou les expressions qui l'intéressent. Et enfin il indique la présentation qu'il souhaite, et qui peut prendre la forme d'un index, d'une liste de fréquences ou d'une concordance avec un contexte de la longueur souhaitée. Les résultats sont fournis en quelques secondes. On peut les imprimer parallèlement ou en différé, les enregistrer, les transmettre à un autre logiciel "*

A la suite de Frantext, plusieurs banques de données textuelles ont vu le jour un peu partout dans le monde : Quebectext, Beltext, Suitext, plus proche de le projet de Ibn Rushd de l'université marocaine.

Nous envisageons la continuité de ce projet, la BnTA, en complétant la base par des liens hypertextuels tels les informations para textuelles, les types de texte, la chronologie, la critique universitaire et médiatique, les éditions renouvelées etc.

### **Bibliographie**

Brunet E., "*Apport des technologies modernes à l'histoire littéraire*" in *L'histoire littéraire aujourd'hui* (direction Béhar Henri et Fayolle Roger), Paris : Armand Colin, 1990, p.99.

Charaudeau P., Maingueneau D., *Dictionnaire d'analyse du discours*, Paris, Seuil, 2002.

Joubert J-L., "*Le réseau "littératures francophones" de l'UREF et la recherche bibliographique*" in *Les banques de données littéraires comparatistes et francophones*, textes réunis par Alain Vuillemin, Limoges : Pulim, 1992, p.30

Moisan C., *Qu'est-ce que l'histoire littéraire ?*, Paris : PUF (coll. "Littératures modernes"), 1990.

Revue Informatique et Statistique dans les Sciences humaines, XXXII, 1 à 4, 1996.  
C.I.P.L. - Université de Liège -