La recherche plein texte

L'utilisation du CD-Rom de base de données textuelles DISCOTEXT1

par Pierre-Marc de Biasi
(ITEM-CNRS)

Abstract : Le CD-rom DISCOTEXT 1 coédité par CNRS/Hachette est une base de données textuelles regroupant 300 œuvres littéraires et un logiciel d'analyse. Le corpus complet offre un choix représentatif de la production littéraire française entre 1827 et 1923. C'est sur cet ensemble de 42 millions de mots-occurrences que le logiciel d'analyse propose un mode d'investigation originale. Disposant de nombreux critères et outils pour définir son corpus et sa question, le chercheur peut demander une recherche d'occurrences ou de cooccurrences paramétrées à deux ou trois termes. Le logiciel permet d'accomplir, sans risque d'erreur, en quelques minutes, des recherches plein texte demandant des mois de travail par une méthode traditionnelle.

DISCOTEXT 1[1] a été conçu pour permettre une gamme très étendue de recherches de type "plein texte" dans une partie de la base de données textuelles FRANTEXT. Ce CD-rom compile beaucoup plus qu'un florilège : on y trouve 108 auteurs et un peu plus de 300 œuvres représentatives de la littérature française entre 1827 et 1923, c'est à dire l'équivalent d'environ 7 mètres-linéaires de bibliothèque, l'ensemble étant structuré comme un tout accessible transversalement en chacun de ses éléments (les mots-occurrences), et rendu lisible (ou, en tout cas, explorable) à une vitesse électronique, par un brillant système d'analyse textuelle. C'est ce gigantesque corpus de 42 millions de mots-occurrences qui s'ouvre aujourd'hui à un nouveau style d'investigation : grâce à DISCOTEXT 1, à la souplesse de ses instruments d'analyse, il devient possible de parcourir, à loisir et dans leurs moindres détails, des milliers de pages, de retrouver ou d'identifier en quelques secondes une citation, de se constituer en quelques minutes, de précieux fichiers de références qui, par les méthodes traditionnelles, auraient demandé des dizaines d'heures de travail, de découvrir des centaines d'occurrences dont le rassemblement et le tri auraient exigé des mois de lecture... Bref, il s'agit d'un bel instrument d'investigation, qui risque bien de transformer radicalement les conditions de la recherche en littérature. Ce produit, relativement nouveau, mais encore assez peu diffusé, mérite d'être connu et utilisé par le plus grand nombre. Sans dissimuler ses limites, on essaiera ici de décrire son utilisation -à vrai dire très simple- et quelques unes de ses performances souvent étonnantes.

Etendue et limites de la base

Le secteur de FRANTEXT disponible dans la base DISCOTEXT 1 contient plus de 300 oeuvres, ce qui signifie que, même avec ses dizaines de milliers de pages, cette compilation n'est qu'une sélection, représentative certes, mais restreinte. A quels critères

---

[1] Le CD-rom *DISCOTEXT 1* est une coproduction CNRS/INaLF - Hachette Supérieur - Bureau van Dijk, publiée en co-édition CNRS-Hachette, avec le concours du ministère de la recherche et de la technologie (DIST). Il contient des extraits de la base de données textuelles élaborée par l'Institut National de la Langue Française (INaLF) du CNRS et un logiciel d'analyse réalisé par le Bureau Van Dijk avec la collaboration de J. Dandien (INaLF). Pour les éditions Hachette, le produit a été finalisé et publié sous la direction d'Alain Pierrot. Ce CD-rom, utilisable sur PC, est diffusé par la maison Hachette. Il se présente sous la forme d'un coffret contenant un disque CD-rom, un guide d'utilisation et un fascicule de sources bibliographiques.

cette sélection a-t-elle obéi? et comment se présente l'état de la base complète? Les choix, les proportions et certaines lacunes peuvent étonner; mais une sélection est toujours discutable, cela va sans dire, et le produit n'aurait pas pu exister sans cette sélection. Bornons-nous à un rapide tour d'horizon. Cent-huit auteurs sont représentés par un peu plus de trois-cents œuvres, ce qui ne donne pas du tout une moyenne de trois œuvres par auteur. Sur ces 108 écrivains, 50 ne sont présents que par 1 œuvre tandis qu'une bonne quinzaine d'autres atteignent 7 à 8 œuvres ou plus, certains s'approchant d'une situation du type œuvres complètes. Ce choix se défend parfaitement : l'immense avantage de cette sélection qui parvient à maintenir une hiérarchie forte (les "grands" écrivains sont en général bien représentés), c'est qu'elle permet en même temps d'intégrer (notamment par les auteurs à un texte) toute un univers de la production littéraire qui échappe souvent à la recherche : des échantillons représentatifs de la littérature à succès de l'époque dans tous les genres, parmi lesquels des œuvres rares devenues inaccessibles aujourd'hui, des textes de scientifiques (Ampère, Claude Bernard, etc.), d'historiens (Fustel de Coulanges, Tocqueville, etc.), de philosophes (Comte, Cournot, Renan), de sociologues (Durkheim), de penseurs politiques (Jaurès, Leroux, Proudhon, Sorel), un florilège de littérature intime, et surtout une foison de correspondances, journaux (Goncourt, Guérin, Michelet, Renard, etc.), cahiers et mémoires qui constituent un véritable trésor d'autant plus précieux que le système rend, pour la première fois dans leur histoire, tous ces textes de part en part transparents à l'utilisation.
Cette richesse, en général, ne s'est pas gagnée aux dépens des œuvres majeures qui sont solidement installées dans la base. L'œuvre de Balzac est représentée par dix titres importants. Les flaubertiens n'ont pas à se plaindre : la base contient toutes les œuvres de la maturité de *Madame Bovary* à *Bouvard et Pécuchet*, avec en prime, *Souvenirs notes et pensées intimes*, et de très larges extraits de la *Correspondance* : les lettres de 1848 à 1880. Stendhal, en revanche, n'est pas bien servi : *Le Rouge et le Noir*, *Lucien Leuwen*, et *L'Abbesse de Castro* : c'est tout; ça fait un peu mince. Chateaubriand n'est présent que par *Les Mémoires d'Outre-Tombe*, mais avec un texte intégral. Des Goncourt, il a été retenu *Charles Demailly*, *René Maurepin* et *Madame Gervaisais*, ce qui n'est déjà pas si mal, mais les dix-neuviémistes sauteront de joie en sachant qu'ils disposent de la totalité du *Journal : mémoire de la vie littéraire* (soit de l'équivalent de plus de 5000 pages dans l'édition Ricatte de 1959). Maupassant totalise plus de 3700 pages de *Contes et Nouvelles*; Zola est sans doute le mieux loti puisqu'il compte à lui seul 21 œuvres : on y trouve *Thérèse Raquin* et *Madeleine Férat*, et tous les romans du cycle des Rougon-Macquart à l'exception du 8ème volume —*Une Page d'amour*, 1878— qui est certes un roman un peu marginal, mais dont l'absence sera regrettée par les chercheurs qui auraient voulu travailler sur un état complet de cet important corpus. Huysmans est richement représenté par 7 œuvres; même score pour pour Maurice Barrès, une de plus pour Anatole France. Les proustiens seront ravis : ils pourront travailler plein texte sur l'intégrale de *A la recherche du Temps perdu* (dans l'édition P. Clarac et A. Ferré à la Pléiade, 1961). La base privilégie visiblement les œuvres de prose narrative, mais l'essentiel du théâtre est là et la poésie n'a pas été négligée : Baudelaire sans être complet, est richement doté, comme Hugo (mais pourquoi n'y-a-t-il pas *les Châtiments*?), Laforgue, Lamartine, Mallarmé, Maureas, Musset, Nerval, Rimbaud, Verhaeren, Verlaine et Vigny, sans parler d'une belle collection de poètes "mineurs" qui réservent sûrement bien des surprises.

La recherche plein texte sous DISCOTEXT 1

Le principe général d'utilisation de ce logiciel repose sur **trois opérations** : **la création d'un corpus** de travail, **la spécification de la requête**, **la recherche automatique et le traitement des résultats**. Il faut y ajouter l'usage d'outils complémentaires proposés dans les "services annexes" du logiciel. La plupart des manipulations exigées de l'utilisateur restent très proches des conditions traditionnelles de la recherche sur les textes et n'exige aucun apprentissage particulier, même si la base et le logiciel ne travaillent pas à

proprement parler avec des mots mais avec des graphies alphabétiques ou non alphabétiques.

Opération 1 : la création d'un corpus de travail. Pour permettre au chercheur de délimiter aussi précisément que possible l'ensemble des textes sur lesquels il désire travailler, DISCOTEXT 1 propose une série de six critères bibliographiques. En choisissant un ou plusieurs de ces critères, en les combinant, en procédant par élargissement ou restriction, l'utilisateur délimite le champ d'investigation conforme à ses besoins : il pourra, par exemple, créer le corpus des textes de prose romanesque écrits entre 1830 et 1880 qui contiennent le mot "histoire" dans leur titre, à l'exception des oeuvres de Balzac. Une fois la sélection terminée, le corpus créé est immédiatement présenté par le système. L'utilisateur peut aussitôt en **visualiser** les références, **effectuer une seconde sélection** plus fine, **trier** ces références, les **sauvegarder** dans un fichier-corpus, les **éditer**, les **imprimer**, et **combiner** ce corpus avec un autre fichier-corpus déjà créé. Lorsque le corpus de travail correspond exactement aux objectifs de la recherche, le chercheur peut passer à la seconde opération : la spécification de sa requête.

Opération 2 : la spécification de la requête. Spécifier l'objet de la recherche consiste à indiquer au système le ou les élément(s) que l'on souhaite localiser dans le corpus de travail. L'étude portera aussi bien sur un mot isolé que sur un syntagme, une séquence de mots, ou une citation complète. Le chercheur pourra également utiliser des listes de mots, par exemple en regroupant automatiquement les formes conjuguées d'un verbe, ou en formant sur mesure un ensemble de mots se rapportant à un thème, etc. Pour mettre au point sa requête, l'utilisateur dispose d'outils (*troncature, paramétrage*) qui enrichissent la recherche de cooccurrences en permettant par exemple d'imposer un ordre d'apparition des éléments. Une fois que cette opération de spécification est terminée, il ne reste plus qu'à lancer la recherche automatique pour recueillir les résultats.

Opération 3 : la recherche automatique et le traitement des résultats. Après quelques instants d'exploration automatique, DISCOTEXT 1 propose la **lecture des résultats** de sa recherche : le chercheur peut **visualiser**, les uns après les autres, les extraits de textes contenant les occurrences trouvées, et utiliser au besoin un **zoom** pour élargir sa lecture. Il lui sera également possible d'opérer une **sélection** dans les résultats et de supprimer les occurrences qui lui seraient inutiles, de constituer un **fichier-résultat**, de le **trier** alphabétiquement ou chronologiquement, et, bien sûr, de l'**éditer** et de l'**imprimer**.

Conditions de l'utilisation : touches, signes conventionnels et procédures. L'utilisation de DISCOTEXT 1 n'exige aucune formation particulière en informatique : on peut commencer à travailler utilement dans l'heure qui suit la première prise de contact avec le système. En pratique, l'utilisateur n'est conduit à manipuler que des mots ou des textes écrits en langue française, en procédant comme il le ferait traditionnellement avec des fiches et des carnets de notes. La seule différence est que ces supports papier seront ici remplacés par des fichiers informatiques, beaucoup plus simples à gérer en réalité, et que le travail de lecture et de recherche dans le corpus est effectué automatiquement par le système. Il reste néanmoins que le chercheur devra faire des choix, donner des ordres au système et lui fournir les éléments textuels sur lesquels il souhaite le faire travailler. Pour ces manipulations très simples, l'utilisateur est amené à utiliser quelques signes conventionnels, et des fonctions correspondant à certaines touches du clavier; mais leur usage ne rencontre réellement aucune difficulté de compréhension ni de mise en œuvre.

***

Création d'un corpus de travail :
délimiter le champ de la recherche

Le "corpus de travail" est l'ensemble des oeuvres sur lesquelles le chercheur entend effectuer ses recherches. Au démarrage de l'application, le système considère par défaut que le corpus actif s'étend à la totalité des oeuvres contenues dans la base, à savoir plus de 300 oeuvres, enregistrée en 579 textes (les oeuvres longues, ou publiées en plusieurs tomes, ou rédigées à des dates différentes, etc. ont été mises en mémoire sous la forme de plusieurs "textes" liés). En début de session, la mention "base complète" apparaît donc en haut de l'écran à droite, sous l'indication "Corpus actif : 579 textes". La création d'un corpus de travail consiste à sélectionner un certain nombre d'oeuvres par une recherche de type bibliographique. A cet effet, le système met à la disposition de l'utilisateur une série de six critères de base : ***le nom de l'auteur***, ***le titre de l'oeuvre***, ***les mots du titre***, ***la date de l'oeuvre*** (date présumée de fin de rédaction, ou de la première édition, ou de la première représentation pour les oeuvres théâtrales), ***le genre*** **de l'oeuvre**, ***la forme*** **: prose ou vers**. En choisissant un ou plusieurs de ces critères, en les combinant (par les fonctions logiques OU, ET, SAUF), en procédant par extension ou restriction, il devient possible de délimiter un champ de recherche. Une fois précisées les limites de ce champ, DISCOTEXT 1 cherchera dans la base tous les textes correspondant à la demande et les "activera", c'est à dire en fera le corpus spécifique et exclusif de l'investigation. Le corpus créé est aussitôt présenté à l'utilisateur pour qu'il puisse **visualiser** les références, **effectuer une seconde sélection** plus fine en ne retenant que les textes qui l'intéressent vraiment, **trier** ces références, les **sauvegarder** dans un fichier-corpus, les **éditer**, les **imprimer**, et , le cas échéant, **combiner** ce corpus avec un autre fichier-corpus qu'il aurait déjà créé.

Restreindre, élargir, exclure. Le Menu général de création d'un corpus se compose de 7 options donnant accès à différents sous-menus de travail. Les trois premières options : 1-***Restreindre le corpus (ET logique)***, 2-***Elargir le corpus (OU logique)***, 3-***Exclure un critère (SAUF logique)*** servent à définir le corpus, et conduisent toutes les trois au menu des critères bibliographiques (**Nom d'auteur**, Titre, Mots **du titre** , **Prose ou vers**, Genre, et **Date**). Mais, bien entendu, ces critères bibliographiques ne délimiteront pas le corpus de la même manière selon que l'on aura opté pour *Restreindre*, *Elargir* ou *Exclure*. Prenons l'exemple d'une recherche qui devrait porter sur le corpus des oeuvres en prose, écrites ou publiées dans la période 1840-1880, dont le titre comporte le mot "histoire" ou le mot "vie", excepté les oeuvres appartenant au genre *Mémoires*. Au moment où il s'agit de créer le corpus, le système commence par donner accès à la "base complète". L'option *Elargir (OU logique)* n'est donc pas initialement active; elle le deviendra dès que le chercheur aura délimité un premier ensemble de textes plus restreint que la base complète. Dans l'exemple choisi, la procédure sera la suivante :
a./ le chercheur sélectionnera d'abord l'option *Restreindre (ET logique)* pour restreindre la base au sous-ensemble des textes dont le titre comporte le mot "histoire" (par le critère "mots du titre" : histoire)
b/ puis il se placera dans l'option *Elargir (OU logique)* pour sélectionner également les textes dont le titre contient le mot "vie" (toujours par le critère "mots du titre" : vie)
c/ ensuite, il devra revenir à l'option *Restreindre (ET logique)* pour ne sélectionner que les textes en prose (critère "prose ou vers" : prose). Il disposera alors du corpus des textes en prose dont le titre contient "histoire" ou "vie". Il ne lui restera plus qu'à délimiter la période et à exclure le genre Mémoires.
d/ il lui suffira d'utiliser l'option *Restreindre (ET logique)* pour délimiter la période 1840-1880 (par le critère "date" : période entre 1840 et 1880)
e/ puis de sélectionner l'option *Exclure (SAUF logique)* pour rejeter de son corpus de travail tous les textes qui relèveraient du genre "Mémoires" (par le critère "genre" : Mémoires).

Cette utilisation des fonctions ET, OU, SAUF ne présente aucune difficulté particulière, mais oblige à une certaine vigilance logique qui peut être considérée comme plutôt saine intellectuellement. Les risques de mauvaise manœuvre portent surtout sur le choix entre

ET et OU. Dans l'exemple que l'on vient de voir, si l'utilisateur avait utilisé l'option *ET logique* dans l'étape b, pour demander la présence du mot "vie" dans le titre , il aurait en fait restreint son corpus aux textes dont le titre contient à la fois "histoire" et "vie" . Si, au contraire, il avait utilisé l'option *OU logique,* dans l'étape d, pour préciser la période chronologique 1840-1880, il aurait agi comme s'il souhaitait ajouter à son corpus tous les textes de cette période . Cela reviendrait à demander que le système sélectionne ceci (les oeuvres dont le titre contient "histoire" ou "vie"), ou encore cela (tous les textes écrits entre 1840 et 1880, sans se soucier de savoir s'ils comportent ou non "histoire" ou "vie" dans leur titre).

Les critères bibliographiques. Ouverts par les opérateurs logiques, les critères bibliographiques donnent au chercheur les moyens d'une sélection fine des limites de son corpus de travail. Il peut le délimiter par l'*Auteur* en choisissant un ou plusieurs noms dans la liste alphabétique des 108 noms d'auteurs représentés dans la base, ou par le *Titre* de l'œuvre (il y a en plus de 300), ou même par un ou plusieurs mots contenu(s) dans un ou plusieurs titre(s). Le chercheur a aussi faculté de préciser son choix par les critère de forme (Prose/Vers) et de Genre (*Correspondance, Eloquence, Essai, Mémoires, Poésie, Presse, Récit de voyage, Roman, Théâtre,* ***Traité***). Dans le détail, quelques appartenances génériques pourraient être discutées, mais globalement, les répartitions sont fiables et fonctionnent parfaitement. Enfin, outre ces cinq critères bibliographiques, utilisables de manière croisée, le chercheur a encore la possibilité, évidemment très précieuse, d'affiner son choix par le critère *Date* qui lui donne accès à quatre options chronologiques utilisables dans les limites du corpus (entre 1827 et 1923, ces années étant incluses) pour sélectionner les oeuvres par leur date (date présumée de fin de rédaction, de première édition, ou de première représentation pour les oeuvres théâtrales). L'option *année* permet de sélectionner les oeuvres datées de l'année qu'il suffit alors de spécifier. L'option *Postérieure à* sert à sélectionner les oeuvres postérieures à l'année spécifiée (cette année étant incluse) : si, par exemple, le chercheur tape 1887, le système sélectionne les textes datant de la période 1887-1923. L'option *Antérieure à* permet de sélectionner les oeuvres antérieures à l'année spécifiée : 1887 sélectionne les textes datant de la période 1827-1887. Enfin, l'option *Entre* permet de sélectionner les oeuvres datées de la période comprise entre deux années spécifiées (ces deux année étant incluses) : si l'utilisateur tape 1830-1848, le système sélectionne les textes datant de cette période de 19 années. Ce critère chronologique se combine bien entendu avec tous les autres critères bibliographiques. Au cours de cette délimitation du corpus de travail, un écran "Historique des critères choisis" permet à l'utilisateur de contrôler les sélections qu'il effectue, étape par étape : à chaque nouveau critère retenu (sous la catégorie restreindre, élargir ou exclure), le système affiche le nombre de textes concernés par ce choix.

Visualisation des références. Dès que la délimitation est jugée satisfaisante par l'utilisateur, il lui suffit de valider. Le système affiche aussitôt le nombre de textes sélectionnés ("corpus actif") et donne à lire, dans l'ordre alphabétique des noms d'auteurs, la liste complète des textes sélectionnées. Si cet ordre alphabétique par auteur ne convient pas, le chercheur peut demander un autre classement. Chaque référence est assortie d'une notice bibliographique complète et précédée d'un double numéro indiquant sa situation dans la liste, et le nombre des références du corpus actif (4/13 signifie : 4ème texte du corpus actif comptant au total 13 références).

Sauvegarde du corpus de travail **:** un écran de sauvegarde permet de mettre en mémoire le corpus créé sous la forme d'un fichier qui restera désormais disponible pour de futures recherches. En sauvegardant le corpus, on lui donne un nom, qui sera automatiquement suivi du suffixe **COR** introduit par le logiciel (pour le distinguer des autres types de fichiers) et on peut aussi y joindre un petit commentaire pour l'identifier avec précision : par exemple, tel fichier enregistré sous le titre "HUGO1**.COR**" pourra avoir commentaire : "les œuvres de prose écrites par V. Hugo entre 1830 et 1860, sauf *Le Rhin* ". Ce dispositif

qui précise le contenu du corpus de travail créé est particulièrement utile lorsqu'il s'agit de "reprendre" d'anciens fichiers-corpus.

Tri des références du corpus **:** le fichier du corpus de travail peut faire l'objet d'une présentation personnalisée. Une fonction *Tri* permet de classer les références chronologiquement ou alphabétiquement : *le tri chronologique ascendant* ordonne le corpus créé sur l'axe du temps en reclassant chaque texte par sa date, du texte le plus ancien au texte le plus récent, et *le tri chronologique descendant*, du texte le plus récent au texte le plus ancien; *le tri alphabétique sur le titre* présente le corpus alphabétiquement d'après les titres des textes; *le tri alphabétique sur l'auteur*, d'après les noms des auteurs

*Edition des références***.** L'écran d'édition sert à affiner la sélection dans les références du corpus de travail. Quelle que soit la précision avec laquelle a été élaborée la sélection par critères, il peut advenir que les références proposées soient en nombre trop élevé, ou qu'elles ne correspondent pas toutes exactement à l'étude envisagée, ou encore que l'objectif de la recherche ait lui-même évolué au vu des résultats et que les zones d'investigation se soient déplacées ou resserrées. Il sera donc possible au chercheur, cas par cas, d'éliminer (par désélection) chacune des œuvres qui ne seraient pas utiles, et de ne conserver (par sélection) que les références indispensables à sa recherche.

Combinaison de deux corpus de travail. Le chercheur a la faculté de "reprendre" un fichier-corpus pour en faire à nouveau un corpus de travail actif; il peut aussi reprendre un second corpus pour le combiner avec le premier. Dans ce cas, c'est le système qui proposera automatiquement trois types de combinaison possibles : par *fusion*, *intersection* ou *différenciation*. Admettons par exemple que le chercheur ait repris, comme *corpus 1*, JOURNAL.COR (tous les textes relevant du genre "journal") et, comme *corpus 2*, GONCOURT.COR (tous les textes des frères Goncourt), les combinaisons possibles sont :
Fusion des deux corpus : tous les éléments du *corpus 1* s'ajoutent à tous les éléments du *corpus 2* . Le corpus combiné fera la somme des textes relevant du genre "journal", et des oeuvres des frères Goncourt, qu'elles relèvent ou non du genre "journal"
Intersection des deux corpus **:** ne seront retenus que les textes appartenant à la fois au *corpus 1* et au *corpus 2* . Le corpus combiné se limitera au *Journal* des frères Goncourt, et rejetera tous les autres textes des Goncourt, ainsi que tous les textes relevant du genre "journal" écrits par d'autres auteurs.
Différenciation des deux corpus (1-2) **:** sont retenus tous les éléments du *corpus 1* à l'exception de ceux qui, dans ce *corpus 1,* appartiendraient aussi au *corpus 2* . Le corpus combiné conservera tous les textes relevant du genre "journal" sauf ceux des Goncourt.
4-Différenciation des deux corpus (2-1) **:** sont retenus tous les éléments du *corpus 2* à l'exception de ceux qui, dans ce *corpus 2*, appartiendraient aussi au *corpus 1*. Le corpus combiné conservera tous les textes des frères Goncourt sauf le *Journal.*

Spécification de la requête :
l'objet et la nature de la recherche

Le corpus de travail étant créé et correspondant avec précision aux textes sur lesquels doit porter la recherche, il s'agit maintenant pour le chercheur de formuler les questions qu'il entend poser au système : dans le vocabulaire de DISCOTEXT1, cela s'appelle "spécifier la requête". On accède à cette étape de travail par l'option 4 du Menu général de DISCOTEXT1 : 4-*Explorer les texte***s**, qui conduit au *Menu général de spécification de la requête* par lequel le chercheur va pouvoir définir la nature et l'objet de son investigation dans le corpus créé, puis lancer la recherche proprement dite. En pratique, cette spécification va consister à préciser le ou les élément(s) textuel(s) que l'on souhaite localiser dans le corpus créé. DISCOTEXT 1 demande à cet effet que lui soi(en)t indiqué(s) le (ou les) **mot**(s) ou **forme**(s), ou les **listes de mots**, ou encore les **séquences de mots** (des expressions, des phrases, etc.) dont il s'agit de repérer les occurrences. Ces éléments

textuels peuvent, en effet, être aussi bien des mots isolés (les mots “amour”, “symboles”, “argent”, “république”, “écrivain”, “laïcité”, etc. pour des recherches socio-historiques, lexicographiques, thématiques, lexicologiques...), que des listes spécifiques (vocabulaire des couleurs, noms de personnages, etc.), des syntagmes, des tours, des citations, etc. : c’est au chercheur de définir sa requête selon ses préoccupations propres (linguistiques, historiques, narratologiques, sociologiques, etc.). Pour y parvenir aussi précisément que possible, il dispose de plusieurs outils assez sophistiqués tels que la *troncature* qui permet de dresser automatiquement la liste des formes comportant une chaîne de caractères donnée (par exemple de tous les mots commençant par "anti...", ou de tous ceux qui se terminent en "...isme", etc.) ou le *paramétrage* qui donne la possibilité de spécifier les recherches de cooccurrences : par exemple toutes les occurrences des mots "enfant", "ange" et "démon" situées, dans cet ordre ou dans un autre, dans la même phrase, ou dans des phrases différentes, à moins de 100 mots l'une de l'autre, ou à moins de 10 mots, etc. L'écran de Spécification de la requête présente trois fenêtres actives (champs A, B et C) : dans chacune de ces fenêtres, le chercheur peut introduire un mot isolé, ou une liste de mots qu'il a créée, ou une séquence de mots, ou encore une séquence de mots et de listes. Disposant de trois champs, le chercheur peut donc procéder à une recherche d'occurrences simples (en utilisant un seul champ, A), ou à une recherche de cooccurrences à deux termes (champs A et B) ou à trois termes (champs A, B et C).

Recherche d’occurrences simples. Si le chercheur se borne à une simple recherche d'occurrences concernant un mot, par exemple les occurrences du mot FLAUBERT dans le corpus JOURNAL des Goncourt, pour retrouver tous les témoignages des Goncourt sur cet auteur, il lui suffit, après avoir créé ou repris ce corpus, d'écrire "flaubert" dans la fenêtre A, et de lancer la recherche en appuyant sur une touche. Il verra aussitôt apparaître sur son écran résultats les passages du *Journal* contenant le nom de Flaubert. C'est simple, mais efficace : en un clin d'œil le chercheur dispose de la série complète des passages utiles, recherche qui, même avec l'aide de l'index, et même pour un œil bien averti, aurait demandé, en lecture traditionnelle, deux bonnes heures de travail. Les avantages du système ne sont que plus impressionnants quand il s'agit d'une recherche d'occurrences portant sur des ouvrages dépourvus d'index ou sur du vocabulaire non indexé : combien de temps faudrait-il pour établir le relevé complet du mot "hérédité" dans le cycle des Rougon-Macquart? Six mois, en travaillant bien? La recherche plein texte vous fourni les résultats en quelques minutes. Et sur le mot "rêve" dans l'œuvre intégral et la Correspondance de Flaubert? Ou encore, sur les mots "jaloux", "amoureux", "chic", "souvenir", etc. dans l'ensemble de *la Recherche du temps perdu*? Pour dire les choses un peu brutalement, DISCOTEXT1 est capable de chercher et de trouver, en une petite journée, le matériel citationnel de base pour cinq à six thèses bien documentées. Ce travail de mise en fiches constituait autrefois le fonds-même de la recherche universitaire. Les résultats obtenus avec DISCOTEXT1 en quelques heures aurait occupé une demi-douzaine de chercheurs pendant de longs mois, et encore non sans risques d'erreurs ou d'oublis, ce qui, évidemment, ne risque pas d'arriver avec l'impitoyable systématicité du logiciel. Mais, bien sûr, les fichiers de résultats ne sont que des fichiers : il reste à les interpréter. Le système ne voudrait pas priver le chercheur de ce plaisir[2].

Recherches de cooccurrences à deux ou trois éléments : La recherche d'occurrence simple, même si elle est longue et pénible, reste envisageable en lecture traditionnelle. Le problème est tout différent pour la recherche en cooccurrence double ou triple. Là, la lecture traditionnelle devient un supplice : imaginez que vous ayez à relever ligne à ligne

[2]Pour se faire une idée des possibilités d'analyse offertes par l'exploration d'occurrences simples, on lira par exemple "Histoire d' "idées reçues" par Anne Herschberg-Pierrot, à paraître dans *Romantisme* 1994 : l'article a été écrit à partir d'une recherche des occurrences de "idées reçues" dans la base FRANTEXT, sur une étendue historique beaucoup plus large (la période 1500-1900) que celle contenue dans DISCOTEXT1, mais la démarche, exemplaire, est parfaitement représentative de ce que peut donner ce type d'investigation .

les cooccurrences proches des mots "curieux", "jaloux" et "amoureux" dans la totalité de l'œuvre romanesque de Proust... Ce travail ne demanderait que quelques minutes à DISCOTEXT1. Mais prenons un exemple plus modeste, dans l'esprit de celui que précédait, sur le témoignage des Goncourt. Admettons une recherche portant non plus sur le seul Flaubert mais sur les relations entre Flaubert et Zola, telles qu'en font état les Goncourt dans leur *Journal* . Il s'agit d'une recherche de cooccurrences à double terme. Sans être impossible, l'investigation serait assez laborieuse en lecture traditionnelle. En recherche plein texte, cela devient un jeu d'enfant : il suffit d'écrire "flaubert" dans la fenêtre A, et "zola" dans la fenêtre B ; et il sera encore possible de raffiner en accédant à un écran de "paramétrage" qui permet, si c'est utile, de préciser la nature du contexte dans lequel devront se situer les deux mots recherchés, leur position relative, et la distance maximale (en nombre de mots) qui peut les séparer. La difficulté ne sera pas plus grande pour une recherche de cooccurrence à trois éléments, par exemple pour obtenir les témoignages des Goncourt portant à la fois sur Flaubert, Zola et Gautier : il suffira d'ajouter "gautier" dans la fenêtre C, en précisant, le cas échéant, les paramètres souhaités .

Recherche portant sur des séquences de mots. Le procédé ne sera pas différent si, au lieu de noms ou de mots isolés, la recherche doit porter sur des séquences de mots : si par exemple il s'agit de retrouver tous les passages du *Journal* où les Goncourt évoquent *L'Education sentimentale* de Flaubert, il suffira d'écrire "l'éducation sentimentale" (la requête est automatiquement enregistrée en minuscules) dans la fenêtre A et de lancer la recherche. Les résultats s'afficheront presque instantanément. Le système est très puissant puisque la séquence de mots recherchée peut compter jusqu'à trois lignes entières d'écran ! On peut également effectuer une recherche portant sur deux ou trois séquences de mots en cooccurrence. Il est également possible de "panacher" la requête et de rechercher en cooccurrence un mot isolé, une liste de mot, et une séquence; ou encore de combiner ces divers éléments textuels en introduisant à l'intérieur d'une séquence, une variable (voire plusieurs) représentée soit par un mot-joker, soit par une liste de mots. Mais, bien entendu, à chaque fois que la requête devra comporter une liste de mots, il faudra que l'utilisateur ait commencé par créer la liste en question.

Création de listes de mots. La possibilité de spécifier une requête en introduisant une liste de mots constitue une ressource extrêmement précieuse pour la recherche. Pour créer une liste de mots, le chercheur dispose de deux procédures : une procédure "manuelle" qui met à sa disposition toutes les formes enregistrées dans la base pour lui permettre de sélectionner un par un les mots qui doivent composer sa liste; et une procédure "automatique" qui met en oeuvre des protocoles provoquant le développement d'une liste par "expansion". L'expansion "conjugaison" permet d'extraire automatiquement de la base toutes les formes conjuguées d'un verbe ; l'expansion "accentuation" donne, à partir d'un mot, toutes les formes accentuées de ce mot qui sont attestées dans la base (c'est-à-dire homographes aux accents près); enfin, l'expansion par "troncature" offre la possibilité d'établir la liste des mots conformes à un modèle donné comportant un certain nombre de caractères fixes, spécifiés, et un certain nombre de caractères indéfinis, variables.

Troncatures. Cette procédure automatique permet de produire, à partir d'un mot tronqué (troncature d'un caractère unique "?", ou d'une chaîne de caractères"*"), la liste des mots présents dans la base qui correspondent au modèle fourni. Si vous tapez par exemple le mot tronqué *?ouche* vous obtiendrez l'affichage d'une liste dont les éléments seront : "*b*ouche, *c*ouche, *d*ouche, *l*ouche, *m*ouche, *s*ouche, *t*ouche" (sans préjuger d'autres graphies, inventées par les auteurs et attestées dans la base . Vous verrez ainsi apparaître : "nouche, rouche, vouche" grâce à Balzac, ...) Cet exemple propose un mot tronqué à l'initiale, mais vous pouvez utiliser le symbole "?" pour la troncature d'un seul caractère à n'importe quel autre endroit du mot. La position d'une chaîne de caractères tronquée,

matérialisée par le signe "*" est laissée à votre libre choix : en début de mot, en fin de mot, ou en position intermédiaire. En écrivant "al*isme", vous produirez une liste dont les éléments, extraits de la base complète, seront : "al*cool*isme, al*exandrin*isme, al*légor*isme, al*pin*isme, al*tru*isme".

Recherche portant sur des ***mots*** *ou des* ***séquences de mots*** : l'écran "Explorer les textes" sert à préciser sur quel(s) élément(s) textuel(s) portera la recherche d'occurrences ou de cooccurrences. On peut introduire, dans chacune des trois fenêtres actives (champs A, B et C), un mot isolé, ou une liste de mots (@), ou une séquence de mots ou encore une séquence combinant mots et listes. Les trois fenêtres sont activées pour recevoir immédiatement la requête : il suffit de taper le texte au clavier, et il s'affiche dans la fenêtre choisie. Soit une requête simple ne contenant par exemple qu'un mot isolé et une séquence de mots : il s'agit d'explorer dans un vaste corpus d'une trentaine d'œuvres la cooccurrence du mot "république" et de l'expression "histoire de France". Une fois le corpus délimité, lorsqu'il sera question de formuler cette requête, il suffira d'écrire "république" dans le champ A, puis de taper "histoire de france" dans le champ B. Si le chercheur ne souhaite pas aller plus loin dans la précision, le système recherchera tous les cas de cooccurrence où ses deux termes se rencontrent à 300 mots de distance maximale l'un de l'autre. Mais le système offre à l'utilisateur la possibilité d'affiner sa requête en la "paramétrant".

Paramétrage. Un écran de travail permet de paramétrer une recherche de cooccurrences en définissant le contexte, l'ordre relatif des éléments, et la distance maximale qui doit les séparer. L'écran "Explorer les textes" comporte trois champs (désignés par A, B et C). Il est donc possible de rechercher trois éléments en cooccurrence (un par champ), et paramétrer, uniformément ou de manière spécifique, les relations AB, BC et AC qui unissent chaque couple d'éléments. Pour reprendre l'exemple cité plus haut, amettons que notre chercheur ait introduit dans le champ A le mot "flaubert", en B "zola" et en C "gautier", et qu'il souhaite trouver les cas de cooccurrences de ces trois noms dans le *Journal* des Goncourt. La rubrique *Contexte*, permettra de préciser s'il entend exiger que ces trois noms se trouvent dans la même phrase, ou dans des phrases différentes, ou si cette variable est négligeable. L'utilisateur pourra aussi traiter les trois éléments spécifiquement par couple : par exemple Flaubert et Zola dans la même phrase, Flaubert et Gautier, ou Gautier et Zola dans un rapport contextuel indifférent. La rubrique *Ordre* servira à spécifier si l'on exige que ces trois noms se présentent dans un ordre précis : par exemple Flaubert avant Zola (A-B : avant), Zola avant Gautier (B-C : avant) ; ou bien dans un ordre quelconque. Enfin, la rubrique *Distance maximale* sert à préciser le nombre maximal de mots qui peuvent être compris entre A et B, B et C, ou A et C : une petite fenêtre permet d'indiquer ce nombre (entre 0 et 300). L'utilisateur peut bien sûr préciser en outre dans quel ordre la recherche textuelle doit s'effectuer : ordre chronologique ascendant ou descendant du corpus actif. En demandant l'ordre ascendant, on obtient d'abord les occurrences attestées dans les textes les plus anciens; avec l'ordre descendant, d'abord les occurrences attestées dans les textes les plus récents.

Recherche portant sur des ***listes de mots***. Le signe @ indique au système qu'il s'agit d'une liste crée par l'utilisateur et non d'un mot isolé. Supposons qu'un chercheur souhaite entreprendre une recherche de grande amplitude (par exemple en base complète) sur les rapprochements littéraires entre la révolution de 1789 et celle de 1830 . Pour cela, il peut par exemple créer la liste "@revol"qui contient le singulier et le pluriel des mots : "agitation, émeute, émotion, événement, insurrection, jacquerie, rébellion, révolte, révolution, sédition, soulèvement, troubles", etc. Le chercheur entend constituer une base de contextes (par exemple des phrases) où l'un de ces mots apparaît en même temps que la séquence "1830" et que la séquence "1789". Il lui suffira de procéder de la manière suivante : -écrire dans la fenêtre A : "@revol" -écrire dans la fenêtre B la séquence

"1830" -écrire dans la fenêtre C la séquence "1789". Puis il paramètre sa recherche en spécifiant que les séquences A-B, A-C et B-C sont dans la même phrase, à une distance quelconque, dans un ordre quelconque. La spécification est terminée, il peut lancer la recherche et recueillir les premiers résultats.

Recherche portant sur ***une séquence de mots, une liste et une*** *séquence de mots et de listes.* Admettons que vous vouliez entreprendre, dans le corpus des oeuvres complètes de Flaubert, une recherche sur la présence d'une expression du type "avec la sensation que l'on éprouve dans les rêves". Vous souhaitez pouvoir retrouver cette expression ou n'importe quelle formulation approchante : qu'elle soit développée comme "avec cette épouvantable et irrésistible impression d'écrasement que l'on éprouve dans les cauchemars", ou brève comme "avec la fluidité que l'on sent dans les songes", etc. Pour y parvenir, il convient de créer trois séquences. Vous avez créé une liste **@sensa** ("*sensation, aisance, impression, agilité, facilité, fluidité, liberté, soulagement, délivrance*", etc.), une liste @eprouv ("*éprouve, ressent, sent*", etc.), et une liste @reve ("*cauchemars, phantasmes, rêves, songes*", etc.). Vous pouvez procéder de la manière suivante : -taper le mot *avec* dans le champ A -introduire la liste *@sensa* dans le champ B -construire dans le champ C une séquence combinant mots et listes qui pourra par exemple s'écrire : que l'on @eprouv dans les @reve
Ce schéma de requête du champ C permettra au système de reconnaître et de localiser aussi bien l'expression "que l'on éprouve dans les rêves" que les expressions "que l'on ressent dans les songes" ou "que l'on sent dans les cauchemars" etc. En séparant *avec* (champ A), *@sensa* (champ B), et *que l'on @eprouv dans les @reve* (champ C), vous vous donnez en outre les moyens de ne pas exclure les formules comportant des éléments intermédiaires : "avec *l'inquiétante* sensation que l'on..." , "avec *cette bizarre* sensation *de bonheur* que l'on..." Il ne vous reste plus alors qu'à paramétrer (contexte : indifférent ; ordre : A avant B, B avant C; distance maximale : A-B 4 mots, B-C 4 mots, A-C indifférent). En un instant le système retrouve et affiche les cooccurrences conformes à votre requête . Il en existe trois : dans *Salammbô*, *L'Education sentimentale* et *Trois Contes*.

Reprise ou combinaison de liste(s). Naturellement, toutes les listes crées peuvent être sauvegardées, reprises, et combinées. Le fichier de mots mis en mémoire est identifié par le nom que lui donne l'utilisateur (nom qui sera automatiquement suivi du suffixe .MOT pour le distinguer des autres types de fichiers) et, éventuellement, par un commentaire qui servira dans l'avenir à en reconnaître le contenu avec précision. L'utilisateur a ainsi la possibilité de se constituer toute une panoplie de listes qui resteront disponibles pour toutes ses recherches. Il lui suffira de les "reprendre". Elles peuvent aussi être combinées pour donner naissance à de nouveaux ensembles, avec les mêmes instruments logiques que ceux qui sont disponibles pour la combinaison des corpus : fusion, intersection, différenciation. Un écran de travail permet de choisir le type de combinaison à effectuer entre les deux listes reprises, dont les titres et volumes sont affichés, pour mémoire, en haut à droite de l'écran.
Soient, par exemple, une *liste 1*, dite @COULEUR1-MOT (bleu, rouge, jaune) et, une *liste 2*, dite @COULEUR2-MOT (ambre, doré, jaune, mordoré, or), les combinaisons possibles sont donc :
-la fusion des deux listes : tous les éléments de la *liste 1* s'ajoutent à tous les éléments de la *liste 2* . La liste combinée contiendra les 7 mots : "ambre, bleu, doré, jaune, mordoré, or, rouge".
-l'ntersection des deux listes : ne seront retenus que les mots appartenant à la fois à la *liste 1* et à la *liste 2* . La liste combinée contiendra 1 mot : "jaune", seul élément commun aux deux listes.
-la différenciation des deux listes (1-2) **:** sont retenus tous les mots de la *liste 1* à l'exception de ceux qui figurent aussi dans la *liste 2* . La liste combinée contiendra 2 mots : "bleu, rouge".

-*la différenciation des deux listes (2-1)* : sont retenus tous les mots de la *liste 2* à l'exception de ceux qui figurent aussi dans *la liste 1* . La liste combinée contiendra 4 mots : "ambre, doré, mordoré, or".

Recherche automatique et traitement des résultats

Le corpus étant clairement délimité et la requête spécifiée il ne reste plus qu'à lancer le système dans son travail d'investigation, en appuyant sur une touche (F2 ou Valide). Après un temps d'exploration initiale (généralement très bref, mais qui dépend évidemment de l'étendue du corpus et de la complexité de la recherche) DISCOTEXT 1 propose la **lecture des** premiers **résultats** de sa recherche sans attendre d'avoir terminé l'exploration totale du corpus. C'est un atout majeur : DISCOTEXT1 a été conçu pour permettre de consulter et de manipuler les premiers résultats au moment même où le logiciel est en pleine recherche. Tout en poursuivant son investigation, le système affiche en permanence, en haut à droite, le nom et le volume du corpus sur lequel travaille le chercheur, et juste en dessous , l'évaluation des résultats en nombre d'occurrences trouvées et en nombre de textes déjà explorés, au stade présent de la recherche . Si ces résultats ne sont pas nuls, l'utilisateur voit ces nombres augmenter au fur et à mesure des avancées du système dans l'exploration des textes, jusqu'à l'affichage des résultats définitifs. Le reste de l'écran est consacré à l'affichage des occurrences. Chaque occurrence est présentée (avec son numéro d'ordre, la référence du texte et la localisation de l'extrait dans l'édition saisie) par un extrait de six lignes de texte avec mise en évidence à l'écran d'au moins un des éléments cherchés. A cette étape du travail, plusieurs fonctions sont à la disposition du chercheur. Pour élargir le contexte de sa lecture, il peut utiliser un *zoom de visualisation sur le texte d'une occurrence* : au lieu des six lignes de l'écran normal de visualisation, il dispose alors d'un écran de quinze lignes affichées dans lequel il pourra faire défiler l'ensemble de l'occurrence, c'est-à-dire un maximum de 300 mots. Le (ou les) élément(s) cherché(s) apparaissent en surbrillance au centre du texte affiché. En dehors du zoom, le chercheur dispose de plusieurs outils de travail : pour sélectionner ou désélectionner une par une, ou en totalité, les occurrences trouvées, pour visualiser une occurrence par appel de son numéro d'ordre, pour procéder à des sondages dans une masse importante de résultats, pour sauvegarder un fichier de résultats, lui donner un nom et lui attribuer un commentaire dans la perspective d'une réutilisation ultérieure. Ces fichiers de résultats peuvent naturellement être plusieurs fois réédités (pour diverses campagnes de sélections), imprimés, triés (chronologiquement et alphabétiquement). Ils peuvent enfin, comme les fichiers de corpus et de listes, donner lieu à des combinaisons par fusion, intersection ou différenciations.

Services annexes : vocabulaire, fréquence, index.

Les trois opérations fondamentales -création du corpus, spécification de la requête, traitement des résultats- offrent les instruments adéquats pour procéder à une gamme très étendue et très diversifiée de recherches (de type sociocritique, poétique, stylistique, narratologique, linguistique, thématique, lexicologique, etc.). Pour un travail plus spécifiquement lié à l'étude du vocabulaire ou des concordances, notamment dans des explorations de grande amplitude, le logiciel propose trois **services annexes** : l'exploration systématique du **vocabulaire** et des **fréquences** dans un corpus, la constitution d'**index**. Le service *Vocabulaire* permet d'établir l'index complet des mots d'un corpus de travail : il fournit, par ordre de fréquence d'utilisation, ou par ordre alphabétique, la liste complète des mots contenus dans le corpus, assortis de leur nombre d'occurrences. Bien évidemment, ce type de recherche automatique demande un certain temps si le corpus est vaste. Le service *Fréquences* permet de demander au système les fréquences d'un mot, ou d'une liste de mots, dans une oeuvre ou dans un corpus. Quant au service *Index*, il constitue pour le chercheur un guide précieux dans la consultation des éditions de référence de son corpus : il pourra en effet demander à DISCOTEXT 1 de lui

donner, pour pour un mot, ou une liste de mots, le nombre total des occurrences et leur localisation (dans l'édition qui a été saisie) : il s'agit en fait d'un véritable index sur mesure portant sur n'importe quel mot de l'oeuvre. Bien entendu, ces trois services annexes sont assortis des mêmes facilités de **visualisation**, de **sauvegarde** et d'**impression** que les autres services de DISCOTEXT 1. Bref, la base offre à son utilisateur la possibilité de se constituer, à la demande, tous les instruments de concordance ou d'index dont il pourrait avoir besoin.

***

On aura compris que l'auteur du présent article est convaincu des immenses qualités de DISCOTEXT1. C'est un remarquable produit qui rendra certainement d'importants services aux chercheurs et qui devrait être à la disposition de tous les étudiants et de tous les universitaires, dans les bibliothèques, les départements de littérature et d'histoire, et sans doute aussi dans les lycées. Cette brève présentation ne donne certainement qu'une faible idée des capacités du produit : la dimension d'un article ne permettait guère d'aller plus loin, et surtout, personne ne saurait encore parfaitement dire tout ce qu'il sera possible de faire et de trouver avec cet outil qui contient à la fois un univers littéraire et une nouvelle manière de le traverser, de le lire, de l'interroger et de le comprendre. Est-ce à dire que DISCOTEXT1 est à l'abri de toute critique? Certes non. On pourrait commencer par critiquer son prix, trop élevé : il coûte un peu plus de 6000F. Mais est-ce vraiment cher? Les comptes sont vite faits : si le CD-rom contient plus de 300 œuvres, cela nous amène à environ 20F par titre, ce qui reste fort compétitif au regard du marché du livre, même en format de poche, et surtout si l'on considère que plusieurs de ces titres sont actuellement introuvables sur le marché. Autre faux défaut : la présentation des textes affichés à l'écran peut décevoir. Ils ne sont pas justifiés à droite, et sont donnés tels qu'ils ont été saisis pour FRANTEXT, au kilomètre. On s'habitue assez vite à cette présentation sauvage, tout comme à la mise en page un peu vieillotte. Et les vrais défauts? Il y en a un qui est énorme : pourquoi ce CD-rom n'est-il utilisable que sur PC, alors qu'une bonne partie des utilisateurs potentiels sont des universitaires, des étudiants et des chercheurs équipés de matériel Apple? Autre problème sérieux : n'était-il pas possible de saisir pour ce produit des éditions de meilleure qualité? Est-il raisonnable, par exemple, de fournir le texte de la *Correspondance* de Flaubert dans l'antique édition Conard de 1923-1954 qui est largement fautive, lacunaire et expurgée? Une ou deux suggestions pour finir : le chercheur est ravi de pouvoir construire des listes de mots sur mesure. Ne pourrait-il pas disposer, quelque part dans sa panoplie d'outils, d'un bon dictionnaire des synonymes attestés dans la base, et pourquoi pas de quelques autres instruments lexicographiques, qui autoriseraient, par exemple, des procédures automatiques d'expansion sémantique et lexicale. Enfin —on a bien le droit de rêver— que donnerait DISCOTEXT si l'on pouvait lui associer les ressources d'une structure hypertextuelle dans laquelle les textes pourraient, par exemple, être dotés d'un riche apparat critique? Il s'agirait sans doute d'un tout autre produit. Alors formulons le souhait à l'envers : cette idée, somme toute géniale, de navigation et d'analyse plein texte qui anime DISCOTEXT1, ne pourrait-elle servir de modèle pour les grandes bases de données hypertextuelles qui alimenteront nos futurs postes de lecture dans les bibliothèques immatérielles de demain?