

Henri Béhar

Université de la Sorbonne Nouvelle (Paris III)

La boîte en valise ou le poste de travail du littéraire

Non, ce n'était pas la *Boîte en valise* où Marcel Duchamp, en 1941, décida de rassembler la quasi totalité de ses oeuvres, mais cela y ressemblait beaucoup. Et pour cause ! Aux débuts de la Troisième République, les instituteurs étaient dotés d'une boîte de matériel qui leur servait à illustrer les célèbres « leçons de choses » et à faire comprendre, de manière concrète, les divers aspects de la mathématique. Sans honte aucune, les cylindres et les cubes y fréquentaient les polyèdres chers à Jarry, pour le plus grand bien des petits enfants confiés à l'Instruction Publique.

A l'ère de l'Education Nationale, du professorat généralisé et du baccalauréat pour 80% d'une classe d'âge, on ne fournit plus le matériel relégué au Musée pédagogique. Les produits s'étant largement diffusés dans le public, et les prix étant relativement modiques, tout le monde, ou presque, possède un instrument de calcul fort perfectionné, aux capacités largement supérieures à ce qu'on peut lui demander dans la vie courante. Les textes réglementaires l'admettent : nul candidat au baccalauréat n'imaginerait de se présenter aux épreuves sans sa calculette, puisqu'ainsi elle se nomme. Je connais un adolescent astucieux - ils le sont tous à cet âge - qui avait programmé sa calculette (autorisée) pour en faire un mini-dictionnaire franco-russe. Qu'aurait pu dire l'administration s'il avait été surpris avec sa diabolique invention ? Que fera-t-elle demain, lorsque se généralisera le livre électronique, sur lequel on pourra lire aussi bien des dictionnaires bilingues que des encyclopédies, les données parvenant à l'élève à la vitesse de l'électricité ? Mon propos n'est pas d'anticiper sur l'éthique de la pédagogie mais plutôt de m'interroger sur les rapports des littéraires avec les outils à leur disposition. La grande majorité des critiques, des professeurs, des étudiants, ignore les services que peuvent leur rendre les machines qui, par ailleurs, ont pénétré leur univers quotidien.

Loin d'être un panégyrique, ni un plaidoyer, ni un procès de ces automates, ces brèves notes voudraient se borner à décrire l'environnement normal du littéraire, en matière de nouvelles technologies en évoquant les investigations permanentes qu'il est en droit d'en attendre, et, pour finir, amorcer une réflexion épistémologique qui nous fait cruellement défaut.

*

* *

Que ce soit à son domicile, au laboratoire de l'équipe de recherche à laquelle il appartient ou même dans la salle de cours, l'enseignant-chercheur devrait posséder la boîte-en-valise comprenant un minimum d'instruments qui constitueront ce qu'on nommera son « poste de travail ». Peu importe le combat que se livrent les géants de l'informatique, les préférences de chacun pour tel ou tel « univers informatique », il n'en demeure pas moins que les outils ci-après, désormais disponibles pour un utilisateur non spécialisé (je veux dire « qui n'a pas suivi d'études en informatique »), doivent figurer dans la panoplie.

A) LE MICRO-ORDINATEUR :

Avec ses logiciels de base, il est bien plus qu'une machine à écrire perfectionnée. Supposons un travail sur l'un des textes figurant au programme de l'Agrégation des Lettres pour 1992. Supposons que le chercheur se procure (nous verrons ci-dessous de quelle façon) le texte informatisé d'*A Rebours*. A l'aide de la fonction **cherche** du traitement de texte, il est possible de repérer, en un clin d'oeil, les contextes du terme "décadence", que voici :

« La décadence de cette ancienne maison avait, sans nul doute, suivi régulièrement son cours ; l'effémination des mâles était allée en s'accroissant ; comme pour achever l'oeuvre des âges, les Des Esseintes marièrent, pendant deux siècles, leurs enfants en»re eux, usant leur reste de vigueur dans les unions consanguines. » (p.2)

« Une partie des rayons plaqués contre les murs de son cabinet, orange et bleu, était exclusivement couverte par des ouvrages latins, par ceux que les intelligences qu'ont domestiquées les déplorables leçons ressassées dans les Sorbonnes désignent sous ce »om générique : "la décadence". » (p.36)

« L'imperfection même lui plaisait, pourvu qu' elle ne fût, ni parasite, ni servile, et peut-être y avait-il une dose de vérité dans sa théorie que l'écrivain subalterne de la décadence, que l'écrivain encore personnel mais incomplet, alambiqué un baume plus irritant, plus apéritif, plus acide, que l'artiste de la même époque, qui est vraiment grand, vraiment parfait. » (p.244-45)

« En effet, la décadence d'une littérature, irréparablement atteinte dans son organisme, affaiblie par l'âge des idées, épuisée par les excès de la syntaxe, sensible seulement aux curiosités qui enfievrent les malades et

« cependant pressée de tout exprimer » son déclin, acharnée à vouloir réparer toutes les omissions de jouissance, à léguer les plus subtils souvenirs de douleur, à son lit de mort, s'était incarnée en Mallarmé, de la façon la plus consommée et la plus exquise. » (p.265)

« Et Des Esseintes sourit, regardant l'un des in-folios ouverts sur son pupitre de chapelle, pensant que le moment viendrait où un érudit préparerait pour la décadence de la langue française, un glossaire pareil à celui dans lequel le savant Du Cange a noté les dernières balbuties, les derniers spasmes, les derniers éclats, de la langue latine râlant de vieillesse au fond des cloîtres. » (p.266).

Avec la fonction **Tri**, il est, notamment, possible d'ordonner une bibliographie dans l'ordre alphabétique des auteurs ou, si l'on préfère, conformément à l'usage, selon la chronologie. Inutile, je suppose, de s'étendre sur l'aide à la rédaction des mémoires et des thèses que présente la fonction **Index**, indispensable pour le repérage des noms propres d'*A Rebours* quand l'édition de référence n'en comporte point. Moins fréquent est le recours à la notion de **Texte caché** qui peut servir à l'étude des variantes d'un manuscrit, avant toute acquisition d'un logiciel plus élaboré.

Outre l'intérêt que présente habituellement un Système de Gestion de Bases de Données (SGBD), du type DBase ou Paradox, dans la gestion quotidienne d'une classe ou d'un organisme quelconque, on verra, ci-après, dans l'article de Michel Bernard, le rôle primordial qu'il assure dans l'élaboration de la BDHL (Base de Données d'Histoire Littéraire), laquelle peut évoluer, selon l'intérêt des utilisateurs, vers une base de données de B.D., je veux dire ici de bandes dessinées, ou encore des traductions et adaptations d'oeuvres littéraires, et s'adapter à n'importe quel domaine artistique. Plus modestement, voici un exemple d'usage personnel du SGBD. Après avoir établi le texte du dernier volume des *Oeuvres complètes* de Tristan Tzara (Éditions Flammarion, 1991), consacré à ce qu'il dénommait *Le Secret de Villon*, reposant sur la découverte d'un grand nombre d'anagrammes contenues dans les vers du *Lais*, du *Testament*, mais aussi de *L'Embuche Vaillant* (que Tzara lui attribue gaillardement), il me fallait savoir, pour étayer mes notes et sans omission possible, tous les noms propres ou les éléments phrastiques découverts par l'auteur de *L'Homme approximatif*; vérifier que les 320 vers du *Lais* avaient bien tous été décryptés (parfois jusqu'à huit reprises) et à quel endroit de l'étude ils se trouvaient analysés; et, lors des corrections des épreuves, retrouver aussi bien le texte de Villon que celui de Tzara. Autant dire que je n'y serais pas parvenu sans un outil permettant ces classements de tous ordres, dont voici un exemple :

a) Les anagrammes de Villon selon leur ordre d'apparition dans le texte de Tzara :

1.	QUI EST RAMPLY SUR LE CHANTIER	ITIIERS MARCHANT	T.199
2.	QUI EST RAMPLY SUR LE CHANTIER	ITIER MARCHANT	T.199
3.	QUI EST RAMPLY SUR LE CHANTIER	CATHERINE	T.199
4.	SE CELLE QUE JADIS SERVOIE	DE VAUSELLES	T.673
5.	ME PRINST LE VOULOIR DE BRISIER	NOE JOLIS	L.14
6.	VOYANT CELLE DEVANT MES YEULX	VAUXELLE	L.18

b) Les anagrammes de Villon selon leur nom crypté :

CORNU A QUEUTE	POUR CUIDER VOSTRE AMOUR CONQUERRE	E.V.65
CORNU QUEUTE	TOUS JOURS VOUS FONT QUELQUE BEAU CONTE	E.V.58
CORNU QUEUTE	OR BEUVEZ FORT TANT QUE EN PEUT COURIR	T.963
DE VAUSELLE	A TOUS LES DIEULX VENERIEULX	L.23
DE VAUSELLES	EN REQUERANT D'ELLE VENGEANCE	L.22
DE VAUSELLES	SE CELLE QUE JADIS SERVOIE	T.673

c) Les anagrammes de Villon dans le *Lais* :

L.001	L'AN QUATRE CENS CINQUANTE SIX	CERMOYE
L.002	JE FRANCOIS VILLON ESCOLLIER	CERMOYE
L.003	CONSIDERANT DE SENS RASSIS	CERMOIES
L.003	CONSIDERANT DE SENS RASSIS	DENISE
L.004	LE FRAIN AUX DENS FRANC AU COLLIER	FRANCOIS
L.004	LE FRAIN AUX DENS FRANC AU COLLIER	SARMOYE...

B) LE LECTEUR DE CD-ROM :

Les compact-disques numériques (CD-ROM) diffusés en France (dont une bonne dizaine concerne actuellement les littéraires) sont consultables sur un ordinateur personnel, à condition de se munir du lecteur adéquat qui peut être intégré à l'appareil ou connecté extérieurement. Quelle que soit la configuration choisie, je ne conçois plus qu'on puisse s'en dispenser pour consulter les atlas, dictionnaires électroniques, encyclopédies en tous genres et autres banques de données gigantesques mises sur le marché, périodiquement renouvelées par abonnement.

Pour illustrer leur intérêt dans l'approche précise des textes, je reprendrai l'exemple d'*A Rebours* tel que l'a traité Hubert de Phalèse (Editions Nizet, 1991). Ce collectif de chercheurs, estimant, à tort ou à raison, que l' énoncé présentait un certain nombre de difficultés pour le commun des lecteurs, a établi un glossaire des mots rares, inusités ou détournés de leur usage courant en s'appuyant sur une « concordance » (ici ramenée à une ou deux phrases avec l'indication de la page dans l'édition de référence) elle-même élaborée automatiquement. Chaque fois que le vérificateur d'orthographe butait sur un terme, on le relevait, puis on contrôlait son sens dans le dictionnaire *Robert électronique* ou *Zyromys* (produit par la firme Hachette) et, toujours automatiquement, on enregistrait les définitions, les explications, les observations les plus pertinentes sur l'usage, comme ici :

affusion : « il supprima les alcools, le café, le thé, but des laitages, recourut à des affusions d'eau froide » (182), « simples affusions froides, suivies d'énergiques frictions » (202). \$ Méd. Vx. Procédé d'hydrothérapie qui consiste à verser l'eau (froide ou chaude) en nappe sur une partie du corps./R

Gondebald : « sous le règne de Gondebald » (167). -\$ Gondebaud, Gondobald ou Gombaud (m. à Genève, 516). Il promulgua la loi gombette (régissant les Burgondes et les Gallo-romains sous un même code, et instaurant le duel judiciaire./Z

Parfois, faute de trouver l'indication voulue dans ces dictionnaires, on poussait la curiosité jusqu'à rechercher une attestation chez d'autres auteurs contemporains, à l'aide de *Discotext1*, CD-ROM élaboré par l'INaLF en collaboration avec Hachette, qui devrait être dans le commerce depuis quelque temps si ne se posaient encore quelques problèmes juridiques. En voici un exemple :

Trautz-Bauzonnet : « il faisait revêtir par Lortic, par Trautz-Bauzonnet [...] d'irréprochables reliures en soie antique » (251). - \$ « c'est le type de l'amateur qui n'a que des livres reliés par Bauzonnet ». Goncourt E. et J., *Journal* (1878), t.2, p. 1094./D

Il ne faudrait pas déduire, des exemples précédents, que le CD-ROM ne sert qu'à des vérifications ponctuelles. En vérité, c'est là son utilisation la plus restreinte, où seule la vitesse d'accès et la maniabilité remplacent l'équivalent traditionnel, c'est-à-dire le support papier. On verra, dans la seconde partie, quelles investigations s'offrent à l'utilisateur de ce type de produit, aussi bien dans le domaine documentaire que dans l'exploration textuelle. D'autant que le compact-disque seulement lisible offre, dans certains cas et pour qui le souhaite, la possibilité d'y adjoindre ses propres données, devenant le CDN (compact-disque numérique) réinscriptible.

J'écarte volontairement du poste de travail, ici envisagé, le CDI (Compact-disque interactif), annoncé pour septembre 1992, qui vise un marché domestique et sera surtout attrayant par les possibilités de travail sur l'image et le son, ainsi que le DDM (Data Disc Man) ou « Livre électronique », mentionné au début et qui, comme le laisse entendre son nom, est surtout un objet baladeur, autosuffisant.

C) ACCES « EN LIGNE » AUX SERVEURS :

Si les CD-ROM ne cessent de se multiplier, ils demeurent coûteux pour l'utilisateur individuel et sont plutôt réservés aux bibliothèques, aux services communs de documentation et aux laboratoires. Cela plaide, d'ailleurs, pour doter le poste de travail du lecteur de disque compact adéquat, l'enseignant-chercheur empruntant à la bibliothèque l'outil de référence dont il a besoin comme il fait pour

un livre. En revanche, il me semble indispensable de munir le micro-ordinateur d'un modem (modulateur-démodulateur) favorisant l'accès aux diverses banques de données, nationales ou internationales, souvent plus complètes que leur version CD-ROM (quand il en existe) et mises à jour de façon permanente. Sans parler du Minitel, produit éminemment national, dont je puis attester qu'il est un outil indispensable pour le chercheur, en particulier le biographe que je suis à mes heures, en me permettant de retrouver tel témoin disparu, j'aborderai, suivant toujours le même exemple d'*A Rebours*, les propriétés de certaines banques de données dont Claire Panigel a, par ailleurs, fourni le descriptif détaillé. Ce sont, évidemment, les banques de données bibliographiques qui intéressent le chercheur, au premier chef. Ainsi *Francis*, banque de données en sciences humaines du CNRS (qui existe aussi en CD-ROM), fournit, outre les références de bibliographie secondaire (essentiellement articles de revues), une analyse du document référencé. Soit l'exemple suivant :

MOURIER-CASILE (P.) : « Modernités à rebours », *Romantisme*, 1983, vol. 13, n° 42, pp. 151-165.

« Modernité fin-de-siècle et Modernité surréaliste. Aussi violemment que les Décadents fin-de-siècle, les Surréalistes rejettent le positivisme scientiste de l'idéologie moderniste. Ils voient dans l'art Décadent un art qui unit à la fois la fin du siècle »nier et le commencement du leur. »

Il m'appartient d'aller lire cet article, dont le contenu correspond à mon besoin. Pour l'heure, j'aurai gagné un temps considérable en apprenant qu'un tel article existe. En outre, chaque référence analysée est précédée de descripteurs, en français et en anglais. Le mode de requête booléenne (qui suppose, malgré sa rigueur logique, un certain apprentissage), m'autorisera à élargir ma recherche à l'ensemble des travaux concernant Décadents et Surréalistes, ou, inversement, à la réduire aux seules études portant sur *A Rebours* et le Naturalisme ou son contraire, l'anti-naturalisme. Dans une telle perspective, la consultation d'une seule source, fût-elle informatisée, ne suffit pas. J'aurai intérêt à interroger d'autres banques de données bibliographiques, à l'étranger, ou encore celles qui consignent plutôt les ouvrages récents, comme *Electre*. Si j'apprends qu'une biographie vient de paraître sur Huysmans, je peux même questionner (sur abonnement) la banque de données du journal *Le Monde* et lire le compte-rendu qui lui a été consacré.

Ce qui vaut pour la documentation bibliographique est, à plus forte raison, pertinent pour les banques de données textuelles dont l'exemple canonique, dans ce pays, est *Frantext*. Inutile de dire, ici, les avantages et les particularités de cette banque. Pour la confection du livre sur Huysmans, qui me sert d'exemple, on a voulu savoir ce que les écrivains pensaient de cet auteur. L'interrogation a fourni deux cent quatre vingt quatorze références, parmi lesquelles on a retenu celles qui traçaient un portrait de l'écrivain et témoignaient de ses rapports avec ses confrères ; puis celles

qui donnaient une idée de la réception d'*A Rebours* par les contemporains. La banque est si riche qu'on risque l'étouffement lors de certaines requêtes. Ainsi le nom de Proust entraîne-t-il plus de huit cents solutions...

D) LA CHAÎNE DE RECONNAISSANCE OPTIQUE DE CARACTÈRES :

Jusqu'ici, j'ai fait comme si le chercheur avait à sa disposition une version numérisée du document sur lequel il travaillait. Supposons qu'il s'agisse d'un texte qui n'est pas encore dans le domaine public, ne figurant pas dans *Frantext*, (à la différence d'*A Rebours*) ou bien d'un corpus formé de plusieurs extraits de presse, par exemple. Si le poste de travail est accompagné d'un scanner et d'un logiciel de reconnaissance des caractères, tel qu'OMNIPAGE ou READSTAR, rien de plus facile que de se livrer à la lecture optique du texte. Comme l'alcool absolu ou anhydre, qui ne dépasse jamais 95 degrés, aucun logiciel n'atteint une fiabilité totale, de sorte qu'il faut effectuer une vérification minutieuse de la saisie. Disons que, pour un livre d'environ deux cents pages au format de poche, une journée de travail/homme est nécessaire. C'est dire qu'il est important d'en prévoir un usage qui justifie l'opération. C'est alors, me semble-t-il, qu'interviennent les logiciels d'analyse textuelle. Canadiens, français, américains, fonctionnant sur le standard IBM-PC ou Macintosh, on commence à disposer d'une quantité non négligeable. Tous ont leurs vertus particulières, leurs mérites et leurs défauts. Dans l'ensemble, ils autorisent des traitements lexicométriques, sur lesquels je ne m'attarderai pas, et surtout lexicaux. Ainsi le logiciel PISTES (Pour une Investigation Systématique des TExteS) de Pierre Muller, diffusé par le CNDP (Centre National de Documentation Pédagogique), permet une indexation alphabétique et hiérarchique du corpus, l'étude des contextes d'un ou plusieurs mots et fournit la liste des « formes » (ensemble des occurrences d'un même mot) communes à tous les textes du corpus ainsi que la liste des formes « spécifiques », c'est-à-dire, grossièrement parlant, celles dont l'apparition n'est pas due au hasard. L'auteur vient de soutenir une thèse à l'Université Paris III Sorbonne-Nouvelle sur les *Etudes socialistes* de Jaurès, dont il analyse avec pertinence le vocabulaire et la rhétorique à l'aide de son propre logiciel. C'est dire combien l'outil, mis au point en vue d'applications pédagogiques, peut aussi servir au chercheur ! Son seul défaut est de limiter la partition du corpus à douze unités, ce qui, dans le cas d'*A Rebours*, (divisé en seize chapitres plus une notice), aurait obligé à un regroupement des chapitres plus ou moins justifié et nous a conduits à lui préférer le logiciel LEXICO1 élaboré par André Salem dans le même laboratoire. Mais ce obstacle semble pouvoir être levé dans une version ultérieure qui s'accompagnera, à ce que j'en sais, d'un logiciel de lemmatisation des formes (les regroupant comme les vedettes d'un dictionnaire).

Il est sûr qu'à l'heure actuelle, aucun logiciel d'analyse textuelle ne peut répondre à la demande d'un chercheur ainsi formulée : « relevez toutes les métaphores du texte ». Cependant, moyennant quelques procédures détournées, on peut obtenir

automatiquement un grand nombre de figures reconnues par les traités classiques. De même, on peut, pour une recherche d'ordre thématique, constituer son champ lexical, de préférence en partant des listes alphabétiques du texte, puis en étendant l'investigation aux termes associés, comme l'ont fait les collaborateurs de *Comptes A Rebours* à la recherche de la vérité dans une âme et dans un corps, dans le réservoir des sens, ou enfin dans l'esthétique du roman. Qui ne voit de quel secours est l'informatique pour le traitement des grandes masses de textes ! Soit la leçon classique : « Portrait de Des Esseintes ». L'inventaire automatique fournit 205 occurrences de ce nom, également distribué dans l'ensemble du récit. Un relevé des contextes (limité à trois ou quatre lignes) offre un repérage immédiat, autorisant une caractérisation rapide de l'unique personnage du récit. Certes, cela n'épargne pas, heureusement, le retour systématique au livre, mais cela permet de s'assurer qu'on n'a oublié aucun de ses traits spécifiques, dans leur propre dynamique. On peut objecter qu'il faudrait, pour être exhaustif, relever aussi tous les pronoms désignant ce noble fin de race. Mais, là encore, la pratique prouve qu'une requête restreinte, adaptée à l'objet particulier et bien formulée satisfait le besoin.

*

* *

En détaillant le poste de travail informatique du littéraire, tel qu'il devrait être équipé aujourd'hui, avec un minimum de logiciels courants, j'ai montré quel type de travaux on pouvait effectuer. Cela ne préjuge en rien la démarche critique à laquelle adhère le chercheur. Je considère, toutefois, que c'est, désormais, le point de départ obligé pour toute investigation. Comment peut-on parler, comme Proust, de « la couleur pourpre » de *Sylvie* chez Gérard de Nerval, ou bien, comme Jean-Pierre Richard, de sa « verdure », si l'on n'a pas, au préalable, vérifié les emplois, dans le texte, de ces termes et de leurs corrélats ? Une rapide interrogation de *Frantext* nous assure que « la couleur de velours pourpre et violacée » ne s'y trouve point. Libre au romancier de rêver, d'imaginer, d'interpréter à partir de sa lecture d'enfance, mais le critique se doit de ne rien avancer qui ne soit établi et je dirais même quantifié. En d'autres termes, c'est l'inventaire des formes du texte qui doit fonder sa démarche, lui permettre de construire sa lecture, d'indiquer des voies imprévues, de dresser des parcours inédits. L'ordinateur est, avant tout, un instrument de validation des hypothèses. Outre sa fonction documentaire et informative, il peut avoir une fonction heuristique, grâce, particulièrement, aux deux logiques qu'il met en oeuvre couramment. La recherche booléenne associe des champs d'investigation a priori disparates. La notion d'hypertexte, particulièrement opératoire dans les CD-ROM, nous permet de « feuilleter » un document, introduisant toute la souplesse et la variété de la navigation dans un univers réputé rigide. Reste que le chercheur en littérature doit savoir évoluer entre le « silence » et le « bruit », ces deux extrêmes de l'informatique. Autant d'aspects sur quoi il convient de réfléchir si l'on ne veut se lancer dans des investigations fastidieuses.