



DUCHASTEL, J. (1991). " Pour une méthodologie d'aide à la lecture et à l'écriture ". In *Actes du colloque international 'Les industries de la langue: Perspectives des années 1990'*, tome 2. Montréal: Office de la langue française/Société des traducteurs du Québec, 1991: 583-601.

Pour une méthodologie d'aide à la lecture et à l'écriture

Jules Duchastel

Centre d'Analyse de Texte par Ordinateur

Introduction

Cet atelier sur les logiciels d'aide à la lecture et à la rédaction ne me semble pas correspondre à un domaine d'application aussi net que ce n'est le cas pour les autres ateliers qui renvoient aux métiers de traducteur, de documentaliste, de terminologue ou encore de formateur. Les présentations de notre atelier semblent pourtant indiquer certaines orientations quant aux applications visées par ce thème. D'un côté, il sera question de détecteurs/correcteurs d'erreurs, chacun prenant plus ou moins en compte les dimensions orthographique, grammaticale ou stylistique de l'écriture. Ces logiciels s'inscrivent dans l'activité générale de révision de tout écrit. D'un autre côté, il sera question de didacticiels d'apprentissage du français, donc d'aide à l'acquisition de compétences pour la production et la compréhension de textes. Une seule présentation se concentre sur l'aide à la rédaction dans un contexte de production experte de textes. Aussi pertinentes soient-elles, on ne peut limiter les domaines d'application à ces trois activités : correction d'erreur, apprentissage de la langue, aide à la rédaction technique.

Je m'attarderai, pour ma part, à resituer la problématique de l'aide à l'écriture et à la rédaction dans un contexte plus global. Je poserai d'abord que la lecture ou l'écriture sont parties intégrantes d'un même processus. Cela me permettra de réfléchir sur la nature de l'objet qui est produit ou manipulé par ces actions, c'est-à-dire le texte. Je distinguerai également les différentes formes possibles de lecture/écriture, ainsi que les divers types de lecteurs/rédacteurs. Après avoir proposé deux typologies de ces outils et de leurs niveaux d'usage, je proposerai d'intégrer ces outils dans une approche méthodologique d'aide à la lecture et à l'écriture.

Lecture et écriture : deux moments d'un même processus

Bien qu'il soit pertinent de distinguer les tâches de lecture et d'écriture comme deux activités relativement indépendantes, il me semble nécessaire de rappeler qu'elles sont les deux moments d'un même processus.



Écrire, c'est lire pour réécrire

Lorsque l'on parle d'aide à la rédaction (ou à l'écriture), on pense le plus souvent à une aide à la réécriture. En amont, il existe bien des aides pour la production matérielle (électronique) d'un texte et des aides pour la production de plans ou de schémas d'écriture. Mais la plupart des systèmes travaillent en aval du premier jet. On corrige, on améliore un premier texte, en somme on réécrit. Et pour ce faire, on fait appel à des fonctionnalités de lecture de texte, assistées ou automatiques, qui permettent soit de repérer des erreurs, des ambiguïtés, soit de prendre connaissance de certaines caractérisations lexicales, sémantiques ou logiques du premier texte. On lit pour mieux réécrire.

Il faut aussi mentionner que l'écriture n'implique pas seulement la relecture de son propre texte, mais la lecture d'autres textes de référence (bases de données textuelles). Encore là les fonctionnalités de lecture permettent d'accéder à ces textes et facilitent ainsi l'acte d'écriture.

Lire, c'est réécrire un nouveau texte

Qu'est ce que la lecture si ce n'est l'écriture d'un sous-texte (méta-texte, hyper-texte) à propos de ce premier texte (hypotexte)? La lecture consiste à produire une représentation d'un texte afin d'en extirper un sens possible. Cette lecture produit un nouveau texte fait de traits ou d'extraits du premier. Ainsi, on obtiendra par l'application de procédures informatiques des informations sur l'occurrence ou la cooccurrence de mots ou de catégories apposées au texte, une sélection de segments significatifs, des résumés, etc. Ces données produites par un mécanisme de lecture automatique ou assistée constituent bien un nouveau texte sujet à interprétation.

En somme, on peut dire que les aides à l'écriture nécessitent presque toujours des fonctionnalités de lecture alors que les aides à la lecture font appel à des fonctionnalités d'écriture dans un sens tout à fait différent : comme modules de génération de sous-textes.

Complexité du texte

La problématique de l'aide à la lecture et à l'écriture conduit inévitablement à la question du texte. A priori, le texte ne semble pas poser de difficulté de définition. Pourtant cette notion renvoie à des réalités multiples. En effet, on utilise ce terme unique pour désigner à la fois des objets différents et une sorte de synthèse de ces objets. Le texte renvoie aussi bien à sa forme graphique (proprement matérielle, sur papier ou sur support électronique), à sa forme linguistique, logique, littéraire, etc., au contenu présumé et aux diverses représentations de ce contenu générées par le lecteur (fut-il automatique ou assisté). Mais, le texte est le plus souvent conçu comme l'ensemble indifférencié de ces objets.

Il me semble important de bien distinguer ces diverses acceptions du texte afin de bien situer les niveaux d'aide à la lecture et à l'écriture. J'introduirai maintenant



quelques distinctions qui me permettront de proposer plus loin une typologie des applications en fonction des usagers.

Forme et/ou contenu

Une première distinction concernant le texte renvoie au rapport entre forme et contenu. Aider à produire ou à lire du texte, pour accéder au sens, implique la prise en compte de la matérialité (forme) linguistique, cognitive, pragmatique ou sociale du texte. Le sens est aussi dans la forme et les stratégies afférentes au contenu sont rarement exemptes de stratégies afférentes à la forme. Cette distinction classique permet de définir un principe de classification des procédures d'aide à la lecture ou à l'écriture : d'un côté, des procédures d'aide à la production ou au repérage d'idées ou de connaissances, de l'autre, des procédures orientées vers la correction ou l'étude des formes. Il est bien entendu que cette distinction permet d'indiquer l'aspect dominant des systèmes, puisque souvent ceux-ci doivent tenir compte à la fois de la forme et du contenu.

Le texte est le figement matériel d'un discours dont l'objectif est de signifier. Sa lecture vise à retrouver une signification quelconque. Sa production sert à donner forme à une pensée, à produire de la connaissance ou de l'information, à communiquer un message, à donner sens à propos du monde des objets ou des idées. Aider à l'écriture, c'est aider à produire ou construire du sens à partir de matériaux linguistiques, cognitifs, pragmatiques et plus généralement sociaux. On parlera, dans le contexte informatique, d'*aide à l'idéation*. Aider à la lecture, c'est produire des représentations linguistiques, cognitives, pragmatiques ou sociales d'un texte, afin de générer un sous-texte signifiant. Du point de vue informatique, on parlera d'*aide à l'analyse de données textuelles*.

Cependant, le travail sur la forme peut constituer un objectif relativement indépendant. Ainsi, dans le processus que nous avons qualifié de réécriture, on met l'accent sur l'aide au repérage et à la correction d'erreurs de forme, quelqu'en soit le niveau. On parlera ici de *logiciels de correction* et d'*aide à la rédaction*. Quant à la lecture des textes, cela nous renvoie à l'activité d'analyse et de compréhension des textes aux divers plans linguistiques et cognitifs. On parlera alors de *logiciels d'analyse de textes*.

Niveaux de complexité

J'ai indiqué plus haut que le texte est une notion à la fois polysémique et synthétique. Le texte est le figement du discours, lui-même le produit d'un processus d'une grande complexité. Il ne peut être de science d'objets complexes sans découpage. Comme il a été nécessaire à de Saussure de distinguer la langue de la parole pour fonder le projet d'une linguistique objective, il nous est aussi nécessaire de découper le texte, ou le discours qui lui est sous-jacent, en entités distinctes. Si cela n'était nécessaire au plan théorique, ce le serait à tout le moins sur le plan empirique, ici technologique. Ainsi, le texte peut être représenté - et alors nécessairement découpé - à divers niveaux de matérialité. Ces niveaux ne sont cependant, ni nécessairement ni



toujours, distinguables du point de vue théorique ou technologique et ils sont évoqués afin de mettre en lumière les problèmes liés au traitement de la complexité.

Le premier niveau - le plus évident pour la technologie informatique - représente le texte comme une suite de caractères graphiques. C'est ainsi que le texte est transformé en codes représentant chacune des formes graphiques. Le traitement de texte - première forme d'aide à l'écriture - peut se contenter de cette représentation. La lexicométrie, en autant qu'elle se limite aux mots comme suites de formes graphiques, produira des opérations et des calculs sur ces mêmes représentations.

Le second niveau est celui de la langue. Le texte est, entre autres, une manifestation d'une compétence linguistique. Celle-ci s'exprime à divers niveaux qui correspondent peu ou prou aux découpages de la science linguistique. Les systèmes de correction s'appliquent à la forme des mots, à leur accord, à leur structuration en syntagme ou aux systèmes de substitution. Les systèmes de compréhension, plus ambitieux, tentent de représenter les phrases ou leurs constituants sous leurs aspects syntaxique ou sémantique.

Le troisième niveau est celui de la structuration textuelle. Le texte s'organise au-delà des phrases qui le constituent, en structures d'arguments ou de narration. Les modules de lecture tentent à ce niveau de reconstituer ces structures, pouvant potentiellement fournir de l'aide à la lecture ou à la réécriture.

Le quatrième niveau est celui de la représentation de la connaissance. Le texte est aussi un contenu. La connaissance qu'il contient, se trouve supportée dans les diverses formes évoquées aux niveaux précédents. Les mots sont des supports sémantiques ou conceptuels. La structure syntaxique ou textuelle supportent également la connaissance. Les modules de repérage et d'indexation de la connaissance servent à décrire le contenu des textes. Les analyseurs sémantiques ou cognitifs visent à représenter des structures de connaissance.

Lectures et écritures plurielles

Parler d'aide à la lecture et à l'écriture nous entraîne donc sur le terrain de la complexité de ce qu'est un texte. Mais, plus pragmatiquement, il est possible de découper également le problème du double point de vue de la nature des textes à produire ou à lire et du niveau d'aide requis. En somme, il n'est ni une seule forme de lecture, ni une seule manière d'écrire.

Nature diversifiée des textes à lire ou à écrire

La nature des textes varie énormément. Ceux-ci se définissent, entre autres, par leur architextualité - appartenance à un genre -. Toute aide à la lecture ou à l'écriture doit donc tenir compte de cette appartenance. Les structures argumentatives, narratives, pragmatiques varieront en fonction du genre. Le contenu davantage expressif, référentiel ou métalinguistique variera également. Il ne peut donc y avoir de système d'aide unique.



Niveau d'aide requis

Le niveau d'aide peut grandement différer. L'emploi du vocable *aide* semble indiquer a priori que les modules informatiques assisteront, de quelque manière, le processus humain de lecture et d'écriture. Dans cette première acception, l'accent est mis sur l'utilisation d'outils spécialisés dans un environnement sous contrôle humain. On peut cependant opposer à la notion d'aide, celle de compréhension. Cette dernière renvoie, dans la tradition de l'intelligence artificielle, au projet de simuler des processus complexes qu'on appelle intelligents. Lorsqu'on parle de compréhension automatique - et même de compréhension assistée - de la langue naturelle, on suggère que l'ordinateur est programmé pour accomplir une grande partie du processus de reconnaissance de la signification contenue dans un texte et se substitue plus ou moins au lecteur humain.

J'utiliserai ici la notion d'aide dans son acception la plus large, c'est-à-dire sans contrainte eu égard à la profondeur de la compréhension recherchée. Dans ce contexte, il faut bien voir que le processus d'écriture/lecture peut alors recevoir une plus ou moins grande part d'aide, selon le type de lecteur ou d'auteur et selon les objectifs qui les animent.

Diversité des lecteur/auteurs

L'aide requise dépend avant tout du type de lecteur ou d'auteur. L'*apprenti* devra oeuvrer dans un environnement conçu pour l'apprentissage de compétences linguistiques, mais également logiques et cognitives. Outre l'intégration de fonctions d'apprentissage, cet environnement devra comprendre des modules d'aide, non seulement orientés vers la correction orthographique, grammaticale ou stylistique de pré-textes, mais vers la construction d'idées dans un contexte logique et/ou narratif. L'*usager de bureautique*, quant à lui, devra disposer des mêmes outils de correction et d'idéation et, en ce qui concerne l'information, d'un accès à des bases de données (dictionnaires, thésaurus, documentation).

Le troisième type d'utilisateur est celui qu'on désigne aujourd'hui comme le *professionnel du texte*. Celui-ci est appelé, dans le cadre de son travail, à manipuler, lire et produire des textes. Il existe plusieurs sous-catégories spécialisées de ce professionnel : l'une plus orientée vers la gestion et le traitement de données documentaires, l'autre vers l'assimilation d'informations textuelles, l'autre encore vers la rédaction de documents. Dans tous les cas, il faut imaginer des principes d'intégration des modules d'aide dans le cadre de modèles méthodologiques ou dans des environnements hyper-textuels.

Enfin, les *chercheurs* dans diverses disciplines - allant des sciences visant l'interprétation sur la base de données textuelles (droit, science politique, sociologie, psychologie,...) aux sciences intéressées à la description des données textuelles (linguistique, études littéraires, sciences cognitives,...) - recherchent la formalisation de procédures avec des objectifs de rigueur, de reproductibilité et de productivité.



Les aides à la lecture et à l'écriture doivent ici encore être pensées dans un cadre méthodologique global.

Objectifs diversifiés pour la lecture ou l'écriture

Quelque soit le type de lecteur ou d'auteur, les objectifs recherchés peuvent également varier ou être combinés dans des équilibres différents. Il me semble que ces objectifs peuvent être ramenés à quatre classes d'activité : une *activité normative*, une *activité informative*, une *activité d'idéation* et une *activité analytique*.

L'*activité normative* recherche la conformité avec des modèles normatifs. Ceux-ci peuvent se trouver à divers niveaux. On recherche la correction orthographique et grammaticale. On désire également obtenir une qualité stylistique minimale. On veut s'assurer de la cohérence textuelle. On veut, à la limite, s'inscrire dans les règles de l'art correspondant à un type donné d'archi-textualité.

L'*activité informative* vise à repérer et à traiter l'information contenue dans les textes ou banques de données textuelles. Cette activité donne accès à l'information brute, mais également permet la reconstruction de la connaissance dans des environnements accessibles. Cette information peut servir, par la suite, à la production de nouveaux textes.

L'activité d'idéation consiste à favoriser le processus de production d'idées. Elle nécessite des environnements plus ou moins intégrés - traitements de texte, idéateurs, modules d'accès à l'information, analyseurs,... - favorisant le processus de production de textes.

Ces deux dernières activités utilisent l'ensemble des connaissances et des technologies disponibles. Elles correspondent davantage aux objectifs des professionnels du texte et de plusieurs catégories de chercheurs.

L'activité analytique consiste à définir et valider des modèles de description des données textuelles à des fins de compréhension en profondeur. Cette activité correspond à la recherche plus fondamentale qui vise le développement de modèles linguistiques et cognitifs du texte. Ces modèles, une fois validés, sont réinvestis dans les trois précédentes activités.

Typologie des applications d'aide à la lecture, à la réécriture et à l'écriture

M'appuyant sur les distinctions que je viens d'évoquer, je proposerai une typologie des différentes formes d'aide. Cette classification me semble couvrir l'ensemble des systèmes qui peuvent être désignés sous l'appellation *aide à la lecture et à l'écriture*. Par contre, je ne fournirai que des indications sommaires sur des exemples concrets de ces applications.

AIDE À LA LECTURE / RÉÉCRITURE

Systèmes orientés vers la norme



Définition : Il s'agit de progiciels dont la fonction ultime est de favoriser l'usage normé de la langue.

Types :

Correcteurs :

orthographiques : généralement intégrés dans les traitements de texte, ils fonctionnent sur la base de la proximité morphologique. Certains systèmes proposent des suggestions de correction. Ils comprennent parfois une fonction d'apprentissage.

grammaticaux : sensibles à la syntaxe au niveau des accords surtout.

stylistiques : correction des barbarismes (mots déformés) et solécismes (emploi syntaxique incorrect); mauvaise ponctuation; lisibilité,...

- *Conjugueurs* : identification des différentes formes des mots.

Exemples :

Correcteurs orthographiques : Word, Word Perfect,...

Correcteurs grammaticaux : Hugo, Ortograf +; en anglais : Grammatik III,...

Correcteurs stylistiques : Le Rédacteur, Nathalie; en anglais : Writer's workbench,...

Systemes d'analyse des données textuelles

Définition : Ces systèmes permettent de produire des descriptions des données textuelles afin de faciliter la compréhension ou la recherche d'information.

Types :

Niveau morpho-lexical Bien que rattaché à la description des lexies simples ou complexes, ce niveau n'exclut pas le recours à la syntaxe.

Lemmatisation : réduction des différentes formes lexicales à leur forme canonique.

Catégorisation : attribution de propriétés (syntaxiques, sémantiques,...) aux mots d'un texte.

Indexation : réduction et organisation de l'information en vue du repérage ultérieur.

Terminologie : identification des termes à l'oeuvre dans un texte.

Locution : blocage des locutions ou synapsies.

Niveau phrastique :

Analyse syntaxique : représentation de la structure syntaxiques des phrases d'un texte.

Analyse sémantique : représentation de la structure sémantique des phrases d'un texte.



Niveau textuel

Analyse cognitive : représentation de la connaissance dans un texte.

Analyse argumentative : représentation de la structure argumentative d'un texte.

Analyse narrative : représentation de la structure narrative d'un texte.

Exemples : Nous ne donnons ici que des exemples au niveau morfo-lexical. Il existe des systèmes agissant aux autres niveaux, mais ils ne comportent pas un caractère suffisamment fini pour être facilement utilisables dans des applications.

Lemmatisation : LCMF (Centre d'ATO)

Catégorisation : fonction "propriété" dans SATO (Centre d'ATO).

Indexation : Kwic, Kwoc, SMART, DIALECT, SPIRIT, CTX,...

Terminologie : Termino (Centre d'ATO), Term Tracer, Mercury/Termex,...

Locutions : Termino et Marquelo (Centre d'ATO).

Systemes d'information

Définition : Il s'agit de systèmes orientés vers le repérage et l'extraction de l'information contenues dans les textes.

Types :

Lexiques : liste des mots d'un texte, vocabulaires bilingues ou unilingues,...

Dictionnaires : sur C.D.ROM, comprenant divers types d'information; on inclut les encyclopédies.

Thésaurus : répertoire de termes normalisés et des relations avec vocabulaire non normalisé.

Base de données textuelles : corpus de textes.

Systèmes de gestion de base de données textuelles (SGBD): permettent d'accéder aux textes et aux connaissances qui s'y trouvent.

Hyper-texte : Système permettant de créer des liens entre différentes bases de données (textuelles, mais aussi graphiques ou autres) et de naviguer entre celles-ci afin d'en retirer les informations utiles à la production d'un nouveau texte.

Exemples :

Lexiques bilingues: Termex, Term Tracer,...

Lexiques : produits par des analyseurs de textes : Word Cruncher, SATO



Dictionnaires : Encyclopédie Grolier, Grand Robert électronique, Encyclopédie Hachette (regroupant le Dictionnaire de notre temps, le Dictionnaire des synonymes, l'Atlas pratique),...

Bases de données textuelles : FRANTEXT, de l'INaLF, comprenant 2,500 oeuvres intégrales du répertoire de la littérature française.

Thésaurus : RAMEAU, THESAUPUS,...

SGBD : Édibase (Inform II Microfor), Seconde (Destin Inc.).

AIDE À L'ÉCRITURE

Logiciels d'édition

Définition : Ces logiciels offrent des fonctionnalités de plus en plus nombreuses pour l'écriture, mais aussi pour l'édition d'un texte prêt à imprimer.

Types et exemples :

Traitements de textes : Word, Word Perfect,...

Éditeurs de textes : Page Maker,...

Logiciels d'idéation

Définition : Systèmes visant à favoriser la conception et l'organisation des idées dans le processus d'écriture.

Types :

"Idea Processors" : logiciels permettant de manipuler et d'ordonner des tables de matières.

Aide à la rédaction : environnements complexes, soit pour l'apprentissage de l'écriture, soit pour aider la rédaction de documents dans des domaines spécialisés.

Exemples :

Ideas Processors : MaxThink, ThinkTank, PcOutline, More,...

Aide à la rédaction : Environnement informatique cognitif stratifié pour la planification et la rédaction de documents techniques.

Typologie d'applications ou de développements selon la catégorie d'utilisateur

Après avoir décrit les divers systèmes correspondant aux divers types d'activités reliées à la lecture/réécriture et à l'écriture, je présente au tableau 1 une synthèse de leur usage en fonction de la catégorie d'utilisateur. En abscisse on trouvera donc les divers systèmes selon qu'ils renvoient à la conformité à la norme, à l'analyse ou au traitement de l'information dans le cadre de l'activité de lecture/réécriture et à



l'édition et à l'idéation dans le cadre de l'activité d'écriture. En ordonnée, on désignera les divers types d'utilisateurs : l'utilisateur de bureautique, l'apprenti, le professionnel et le chercheur. On remarquera que la complexité des systèmes utilisés croît avec la progression dans l'échelle des utilisateurs. D'un environnement orienté outils, on progresse vers un environnement ayant de plus en plus recours à des stratégies d'intégration. D'une conception d'aide, on progresse également vers une conception de plus en plus orientée vers la compréhension en profondeur. On constatera enfin que, pour le moment, les aides à l'écriture comportent peu d'items. Même s'il est vrai que les systèmes de lecture servent en même temps à la réécriture, il n'en demeure pas moins qu'il y a un urgent besoin de développement d'ateliers spécialisés d'aide à l'écriture, intégrant des modules existants dans le cadre d'une finalité d'écriture.

Tableau 1

Types d'applications ou de développements selon la catégorie d'utilisateur

Lecture / réécriture **Écriture**

Utilisateur **Norme** **Analyse** **Information** **Édition** **Idéation**

Bureautique Correcteurs Lexiques Trait. text. Idéateurs

- orthographiques Dictionnaires Édition
- grammaticaux Thésaurus
- stylistiques Bases données

Conjuguers textuelles

- Orienté outils -

Apprenti idem idem idem idem

- Outils + environnement d'apprentissage -

Professionnel idem Catégorisation idem + idem Aide à la

indexation Gestion bases rédaction

Lemmatisation - textuelles



Terminologie - connaissances

Locution Repérage

Extraction

Résumé

Hyper-texte

- Outils + méthodologie -

Chercheur | _____ |

Analyseurs linguistiques :

- syntaxiques
- sémantiques

Analyseurs textuels :

- représentation de la connaissance
- structures d'arguments
- structures narratives

- Développement d'outils et de méthodologie -

Pour une méthodologie d'aide à la lecture et à l'écriture

On a pu constater jusqu'à présent que le domaine des aides à la lecture et à l'écriture était extrêmement vaste et risquait d'apparaître à l'utilisateur comme un immense maquis non balisé. Le problème ne se pose pas vraiment pour l'utilisateur de bureautique qui accédera progressivement à des outils spécialisés répondant à des besoins définis. Il se pose plutôt pour les professionnels du texte qui doivent accomplir des tâches d'une beaucoup plus grande complexité sur les textes. Je poserai, en conclusion, le problème à deux niveaux. D'abord, j'indiquerai les deux orientations qui caractérisent la recherche dans le champ du traitement des langues naturelles. Le rapport dialectique, entre elles, est très important pour comprendre les enjeux qui sont derrière le développement des systèmes qui deviennent accessibles à un nombre croissant d'utilisateurs. J'examinerai ensuite la réponse concrète qui est actuellement offerte aux professionnels du texte, confrontés de plus en plus à des masses de données textuelles à investiguer ou à produire.

Outils ou méthodologie



J'ai proposé, au niveau des définitions, que la lecture et l'écriture sont des activités qui se rapportent au texte, dans toute sa complexité, et que la notion d'aide peut correspondre à divers degrés de profondeur dans le processus de compréhension. Dans ce contexte, se pose donc le problème de réconcilier deux approches qui dominent le domaine de la recherche sur le traitement informatique des langues naturelles.

L'une, *fondamentale*, recherche l'adaptation de modèles linguistiques et cognitifs à des contextes informatiques, afin d'accroître la part de compréhension automatisée. L'autre, *appliquée*, vise la mise au point de techniques d'ingénierie pour le traitement, partiellement assisté et partiellement automatisé, de données langagières.

Les objectifs de la première sont, avant tout, des objectifs de connaissance et sont orientés vers la validation de modèles. Ces recherches sont absolument nécessaires à toute application ultérieure puisqu'elles fournissent les modèles à mettre en oeuvre. Cependant, en raison même de la complexité du langage, cette recherche tend à explorer, sans cesse, de nouvelles pistes - actuellement, la recherche s'oriente vers la prise en compte du contexte d'énonciation, l'élaboration de nouveaux modèles de représentation des connaissances, la prise en compte des logiques dites naturelles - et à chercher des modèles théoriques d'intégration. Cette approche concerne avant tout les chercheurs. Ceux-ci doivent soit se situer au coeur de cette recherche, soit intégrer ses découvertes dans des systèmes d'application.

La seconde approche, s'appuyant sur des découvertes de la première, s'emploie à valoriser les données textuelles à partir d'outils ou de modules adaptés aux divers niveaux de complexité du langage. Elle combine donc approche automatique et assistée, choisit la couverture en largeur avant la compréhension en profondeur et favorise l'interaction entre usager et système. Elle est animée par des chercheurs intéressés au développement de méthodologies d'analyse, mais elle est d'abord destinée aux professionnels du texte.

L'interaction entre ces deux approches ne peut qu'être bénéfique, car elle impose à chacune des critères de validation différents. D'un côté, l'approche fondamentale recherche la validation dans la capacité intrinsèque des modèles de rendre compte de la réalité. De l'autre, la validation s'effectue à travers la capacité des systèmes à traiter des données réelles, en temps réel.

Nécessité d'une méthodologie

Je n'insisterai pas sur l'importance de plus en plus grande d'une nouvelle catégorie de professionnels dont une des attributions principales est d'interagir avec le texte. À côté de ce qu'il est convenu d'appeler les langagiers (terminologues et traducteurs), il existe un grand nombre de personnes dans les organisations qui doivent gérer et accéder à de l'information textuelle ou encore produire des textes en interrelation avec d'autres textes. L'informatisation croissante de l'entrée de textes définit dorénavant le cadre de cette interaction avec les textes. Les pages qui précèdent ont montré que les systèmes d'aide à la lecture et à l'écriture ont tendance à se multiplier. Par contre, ces procédures d'aide sont trop souvent partielles et peu standardisées, ce qui rend difficile leur utilisation intégrée dans le processus de la lecture ou de l'écriture. En somme, il existe peu de stratégies d'utilisation de ces systèmes et leur usage est, pour le moment, limité et ponctuel.

J'ai parlé déjà de la nécessité de découper cet objet qu'est le texte afin d'en permettre la connaissance théorique, mais surtout d'en identifier les ressorts, autant pour sa production que



pour sa compréhension. Ce découpage doit nous rendre conscients de la complexité de tout acte de lecture ou d'écriture. C'est pour cette raison qu'il est normal que les applications disponibles sur le marché nous apparaissent à la fois performantes et décevantes. Dans la mesure où l'on veut accomplir certaines tâches routinières et spécialisées - repérage et correction d'erreurs, recherche de mots,...- les systèmes apparaîtront comme étant relativement performants. Mais, dans la mesure où l'on veut accomplir des tâches plus complexes - saisir le sens d'une information en contexte, résumer un texte,...- alors les systèmes nous apparaissent comme étant très limités.

Lire ou écrire un texte implique, qu'on en soit ou non conscient, la mise en oeuvre d'une méthodologie. Le plus souvent, celle-ci passe inaperçue parce qu'elle est intériorisée dans une pratique qui nous est familière. Le fait de recourir à l'aide de l'ordinateur oblige à concevoir le rapport au texte en pièces détachées. Cela étant, il faut pourtant recomposer la complexité de ce rapport à travers ce que j'appelle une méthodologie de lecture et d'écriture, c'est-à-dire qu'il faut bien identifier les diverses composantes de cette activité et définir les relations qui les unissent. Le sentiment d'éparpillement, ressenti par les professionnels du texte, devant les outils qui leur sont proposés découle de l'absence d'une telle méthodologie.

Dans l'état actuel du développement, une technologie prétend répondre à ce besoin. Il s'agit des systèmes hypertextuels. L'idée sous-jacente à cette approche est de permettre la navigation entre différents systèmes d'information textuelle ou de toute autre nature (numérique, iconique,...). Ainsi, il est possible, à partir d'un traitement de texte, d'aller chercher dans une base de données textuelles (généralement structurée) l'information pertinente à la poursuite de l'écriture. L'hypertextualité est, pour le moment, une idée technologique plus que méthodologique. Les développements vont, avant tout, dans le sens de la mise en relation complexe de divers systèmes les uns avec les autres, permettant ainsi le transfert d'informations.

La réflexion sur l'hyper-textualité va cependant beaucoup plus loin. Par exemple, Edward Barrett distingue entre cette première approche classique de l'hypertextualité, comme pur dispositif de navigation entre bases structurées de données textuelles comportant très peu de possibilités d'intervention sur le contenu de ces bases et une approche topographique de la construction sociale de la textualité. Cette perspective constructiviste adopte une position différente de la conception classique, concevant le texte comme pur produit de la subjectivité. Le texte est vu, au contraire, comme une interaction ou une communication sociale. L'hypertextualité doit donc accroître cette interaction de manière à permettre non seulement la navigation physique entre systèmes, mais la possibilité d'interventions multiples dans les divers systèmes.

"After all, a hypertext is fundamentally a linguistic entity that exists to be manipulated, transformed through a series of collaborative acts either between just one user and the original database (that is, the original programmed structure), or among many users performing various operations upon a central core of texts."

Le projet SACAO (Système d'Analyse de Contenu Assistée par Ordinateur) s'inscrit dans cette deuxième perspective d'hypertextualité., avant même l'usage du terme. Projet appuyé sur l'expérience de chercheurs en analyse de textes par ordinateur, il visait d'abord l'intégration systématique de procédures, existantes et en voie de développement, dans un environnement unifié.



Dans la mesure où les besoins d'accès aux textes sont très diversifiés selon la discipline, l'orientation ou les objectifs des chercheurs, il a fallu penser un système comportant le minimum de préconstruits théoriques. Cela signifie que le système manipule des objets et effectue des opérations qui font habituellement l'objet d'un consensus pour les chercheurs ou, alors, offre à l'utilisateur la possibilité de paramétrer les opérations en fonction de sa propre démarche. Pour ce même motif d'ouverture et en raison de l'impossibilité de répondre, de manière automatique, à la complexité des tâches d'analyse des données textuelles, le système a été conçu pour favoriser le maximum d'interactivité entre les programmes et l'utilisateur et donc le contrôle par le chercheur.

Au fondement de cette approche, le projet s'appuie sur une philosophie constructiviste. La lecture et l'écriture sont comprises comme processus de construction sociale de la textualité. Tout texte se construit sur la base d'autres textes. Les idées et les stratégies d'écriture sont partie d'un même ensemble et se répondent. Les professionnels du texte rencontrent cette réalité tous les jours. Ils doivent souvent réagir à des textes et produire d'autres textes en réponse aux premiers. L'architecture du système est donc conçue pour répondre à cette réalité de l'intertextualité.

Par ailleurs, la compréhension comme la production de textes est un processus itératif faisant intervenir de multiples niveaux. Le système reflète cette complexité des niveaux en donnant accès à une pluralité de modules d'analyse et d'exploration. Le système permet même d'accueillir des modèles non exclusivement compatibles favorisant une lecture et une écriture plurielle.

SACAO comporte quatre modules principaux, eux-mêmes pouvant en comprendre plusieurs. Le premier ensemble comporte l'ensemble des fonctionnalités permettant de gérer les données textuelles (édition, archivage, accès,...). Le second comprend des modules de description des données textuelles. Ces modules correspondent aux systèmes, décrits plus haut, comme systèmes d'analyse. Le troisième comprend des modules de repérage et d'extraction de l'information ou de la connaissance. Ces modules s'appuient en général sur les descriptions produites au niveau de l'analyse. Ces modules renvoient à ce que nous avons désigné comme étant des systèmes d'information. Enfin, le quatrième module comporte des fonctionnalités d'analyse de données textuelles, c'est-à-dire des résultats obtenus par l'exploration des descriptions effectuées sur le texte.

L'intérêt de ce système est de proposer une méthodologie d'analyse supportée technologiquement dans un environnement intégré. Il s'adresse avant tout aux professionnels du texte. Plutôt qu'un cadre conceptuel, il leur offre un espace de travail. Il permet un usage à la fois intégré et stratégique des outils aujourd'hui disponibles. Le projet favorise l'intégration en permettant d'articuler plusieurs procédures, en créant des liens informatiques, en stabilisant des implantations dans des environnements de travail et, à la limite, en visant la plus grande portabilité d'un environnement à un autre. Quant à la stratégie de son usage, le système permet à l'utilisateur de choisir les modules qui sont pertinents à son problème, de paramétrer les opérations retenues par lui et de structurer sa démarche d'un point de départ à un point d'arrivée.

Conclusion

On aura donc constaté que le domaine des aides à la lecture et à l'écriture ne peut facilement se définir, sans faire intervenir une double typologie des activités et des usagers. Tâches



complexes et dialectiquement reliées, la lecture et l'écriture nécessitent la mise en oeuvre d'un grand nombre de nos facultés. Cette multiplicité se reflète dans la profusion des solutions informatiques proposées (traitement de textes, correcteurs, dictionnaires, analyseurs,...). Cependant, ces progiciels sont rarement pensés dans un cadre d'intégration. Tant que l'utilisateur ne cherche qu'une aide ponctuelle pour effectuer une tâche spécialisée, il trouve généralement des systèmes adaptés à cette demande. C'est dans la mesure où le même usager requiert une aide globale pour effectuer un ensemble de tâches complexes de lecture et d'écriture que devient urgente leur intégration dans un cadre méthodologique complet. Avant même d'avoir résolu les problèmes théoriques nombreux qui nous empêchent de comprendre ce qui est vraiment à l'oeuvre dans l'acte de compréhension et de production de textes, il est pourtant possible de penser des environnements informatiques permettant les opérations d'une telle démarche planifiée.