

L'AVENIR D'ELMO 2000 ET DE SES MODULES

Denis FOUCAMBERT

"2 000 arrive, ELMO 2 000 est presque là". C'est ainsi que Robert CARON commence son article. Tout est dans le presque pour Denis FOUCAMBERT qui, après avoir fait le point sur l'état d'avancement des différentes parties du logiciel, énumère les voies ouvertes aussi bien pour l'élaboration et l'amélioration des programmes eux-mêmes que pour leurs utilisations pédagogiques.

Une des difficultés d'appréhension du logiciel ELMO 2000 réside dans le fait qu'il est conçu comme un tout ; ce tout étant, par ailleurs, divisible en un certain nombre de parties indépendantes et autonomes.

Mais, avant de parler du futur, arrêtons-nous un moment sur ce mois d'avril 1992.

Le cœur du logiciel bat dans ce qu'on nomme la base. Il y a encore quelques mois, elle comprenait déjà 35 000 entrées correspondant à encore plus de formes possibles, puisque pour un verbe il peut y avoir 39 configurations convenables et 4 pour un nom ou un adjectif. Pour avoir fait fonctionner la base avec plus de 2,2 millions de mots issus d'articles de presse dans le but d'établir les fréquences d'apparition des vocables, nous nous sommes rendu compte des lacunes de cette base, et avons décidé de l'augmenter encore, et dans un même élan, de la recomposer complètement. Pourquoi recréer ce qui semblait seulement incomplet ? Les prototypes dont nous avons déjà parlé dans ces colonnes s'intéressaient avant tout à des aspects syntaxiques des écrits, et laissaient de côté un pan entier de la linguistique : la sémantique. Quelques ébauches de programmes annexes nous ont convaincus qu'on ne pouvait écarter cette voie. Aussi la nouvelle base comporte-t-elle un grand nombre d'informations supplémentaires. À chaque mot, dans la mesure du possible, seront ajoutées les informations suivantes : ses antonymes, ses synonymes, ses champs lexicaux, les mots qui appartiennent à la même "famille", son préfixe, son radical, son suffixe, sa fréquence, son niveau de difficulté orthographique, son degré de polysémie, ses hyperonymes et hyponymes. À toutes ces informations générales s'ajouteront quelques détails plus spécifiques : présence (ou absence) dans différentes listes comme celles de GOUGENHEIM ou de MESNAGER, type des verbes (auxiliaires de mode, verdictifs,...), type d'organismes (logique, de temps, de lieu)... Cette refonte de la base s'accompagnera de la réécriture de l'analyse grammaticale prenant ainsi davantage en compte les aspects lexicaux.

C'est évidemment un travail de longue haleine, tant pour recueillir ces informations que pour les coder de façon compacte. Cette refonte de la base devrait être bouclée avant la fin de l'année scolaire sous la haute main de Marc BENZIMPA.

Il faut insister encore une fois sur l'importance de tout ce travail qui permettra l'accès à ces informations à tout moment, tant au cours de l'écriture qu'à l'occasion des différentes analyses a posteriori.

Autre chantier en cours : *La genèse du texte*. Ce module s'inscrit dans la continuité des recherches sur les manuscrits d'auteurs (cf Écrit-ratures de Claire DOQUET dans ce même numéro). Cet outil exécute les différents comptages et une analyse récapitulative des modifications effectuées au fil du temps lors de l'écriture d'un texte. Vincent ESTRABEAU y met la dernière main et le premier module devrait être opérationnel dans les jours qui viennent.

Le module *Lecture Méthodique* est lui aussi en chantier. L'ayant déjà présenté en détail (A.L. n°37, mars 92, p.24), je n'y reviendrai que pour dire qu'il avance régulièrement sans véritable problème si ce n'est l'attente de la base définitive qui retarde quelque peu l'analyse sémantique. Mais il est certain qu'il sera prêt dès la finition de la base, les analyses ayant déjà été testées avec les prototypes précédents.

Pour ce qui est de la typologie des productions écrites le problème est tout autre. Nous avons déjà effectué un module expérimental qui classait de façon satisfaisante les textes suivant la typologie de J.-P. BRONCKART (voir A.L. n°32, déc.90, pp86-100). Aujourd'hui nous tentons de créer une typologie plus adéquate, plus complète qui couvrira des champs plus fins que ceux déjà étudiés en enlevant la variable "langage oral" de la classification (récit conversationnel), ce qui permet de cerner davantage les différentes façons d'écrire dans la perspective du dialogisme, des situations et des buts de l'activité d'écriture. Dans ce présent numéro, Thierry OPILLARD présente la méthode employée et Jacques BERCHADSKY les présupposés théoriques. Une fois les textes de références choisis et classés rigoureusement, nous essaierons de déterminer les variables linguistiques qui discriminent au mieux les différents types de productions écrites. L'analyse, après avoir été qualitative (choix des types et des textes) deviendra quantitative et statistique (comptages des indices linguistiques, et formules mathématiques de combinaison des indices pour définir le type du texte).

L'AVENIR

Si deux objets semblent donc proches, un autre, celui de la typologie, reste plus lointain. Comme le signale Thierry OPILLARD, pour le module *Typologie* nous en sommes au classement des textes de référence. Par la suite, les ordinateurs vont "parler". Forts de cette centaine de variables, nous allons analyser les textes pour les classer correctement. Un point semble déjà certain : un travail de fourmi devra être fait à propos du choix des traces linguistiques. Comme capital de départ, nous avons les 27 variables de BRONCKART, ainsi que quelques autres de notre crû. Mais il est évident que nous devons intégrer les résultats d'autres travaux (comme ceux de DUCROT) qui approfondissent les listes de connecteurs logiques et spatio-temporels. De même, un travail est nécessaire sur le rapport que ces connecteurs entretiennent avec chaque type d'écrits. Une autre piste, découverte à propos du travail sur la lecture méthodique requiert notre attention, c'est celle d'une typologie des verbes selon différents types d'actes illocutoires (AUSTIN, 1962 : verbes verdictifs, exercitifs, commissifs, expositifs, comportementaux).

En ce qui concerne les progressions thématiques, il faudra examiner comment les thèmes principaux et secondaires sont introduits, portés, par le flux des phrases et des paragraphes.

Comment toutes ces variables s'insèrent dans les différents types de textes que nous avons définis ? On comprend bien qu'il faudra encore du temps pour mener ces travaux. C'est en quelque sorte un laboratoire linguistique qui aboutira à un produit autonome, mais qui doit encore être développé.

Pourtant, il a déjà des applications. Plusieurs indices que calcule le module de *Lecture Méthodique* sont directement issus du travail sur la typologie. Le but n'en est évidemment pas le même, la présentation non plus, mais ils s'inscrivent en plein dans ce que la lecture méthodique veut observer. Pour exemple, l'analyse des différents pronoms personnels dont la fréquence d'apparition sert à discriminer tel ou tel type de textes et qui se retrouvent dans l'observation des indices d'énonciation telle que la recommande la lecture méthodique.

La sortie du module de *Lecture Méthodique* est envisagée pour la rentrée de Septembre. Cette fois, ce n'est plus dans la conception même du produit que se situe l'avenir, mais dans son utilisation. Quelles pourront être les séances de lecture méthodique quand on disposera d'un tel outil informatique ? Des professeurs de lycée testeront ce produit, ce qui permettra de définir les utilisations possibles (qui seront présentées dans le guide d'accompagnement) et la formation qu'exigera l'emploi d'un tel produit.

Mais il ne s'agit là que d'un outil innovant sur une pratique déjà ancienne comme peuvent l'être les autofocus pour les appareils photo. Tel n'est pas le cas du module *Genèse du texte*, qui apporte une nouvelle façon de décrire une activité langagière, non plus a posteriori mais au cœur même de l'action. Avoir soi-même une vue extérieure, pouvoir s'arrêter à tel ou tel moment important du mouvement pour observer les différentes positions de chacune des parties en jeu, pouvoir comparer plusieurs styles, en les analysant précisément tous les sportifs savent quels progrès a apportés la caméra vidéo pour l'entraînement. Telle sera une des possibilités offertes par la *Genèse*. Mais pour que ce travail soit véritablement productif, il nous faut déterminer les moments importants. Rien de tel que l'observation de professionnels et l'analyse de la construction de leurs textes. Nous savons déjà par les travaux de HAY ou de GRESILLON (cf Écrit-ratures de Claire DOQUET dans ce même numéro) que nous avons à nous intéresser aux différents déplacements des blocs de texte ainsi qu'à tous les remords de l'auteur. Mais le logiciel offre une autre possibilité, inconnue jusqu'ici la mémorisation des temps d'attente. Les pistes ont été décrites dans le numéro précédent des AL (n°37 p73) et nous commencerons à les mettre en œuvre dès le début du mois de mai 1992.

On évoque dans cet article l'étude de bon nombre des aspects de la naissance d'un texte, depuis la prise en compte des situations de départ au point final. J'y rajouterai deux éléments. On y parle déjà de l'analyse des processus de cohérence générale, c'est-à-dire de la prise en compte de la situation de départ, de la macroplanification... jusqu'à la microplanification. Elle s'étudie à partir des unités linguistiques de surface que nous savons retrouver dans un écrit. Mais on peut aussi étudier la gestion dans le temps des variables qui caractérisent le type du texte qu'a voulu produire l'écrivain. Si l'on connaît à l'avance le type du texte, et on le connaît en analysant le produit fini par le module *Typologie*, on peut regarder la façon qu'a eue l'auteur de mettre en œuvre les variables qui le caractérisent. Il s'agit d'observer comment l'écrivain gère la mise en mots du contrat de départ, à quel moment il s'en préoccupe ou si elle est intégrée dès le départ et ne nécessite pas de corrections spécifiques.

En outre, il ne me semble pas avoir vu la possibilité d'étudier, dans la production du texte, la gestion de telle ou telle idée (du "message que veut faire passer l'auteur"), laquelle serait éventuellement repérable par l'étude d'un champ sémantique bien particulier et donc par le repérage de mots s'y référant ainsi que par l'étude des anaphores de ces mots, ainsi que par le choix à tel ou tel endroit d'hyperonyme à la place d'un terme plus précis. En quelque sorte, regarder comment sont gérées, au fil de l'écriture, les quelques idées maîtresses d'un texte.

Toutes ces pistes sont maintenant ouvertes grâce à cet outil. Déjà quelques utilisations en ont été faites (cf. l'article de Thierry HUG dans ce présent numéro), mais l'essentiel est devant. Les versions futures intégreront ce que nous aurons défriché. Tel qu'il existe aujourd'hui, il permet déjà ce que beaucoup de linguistes ont rêvé d'avoir entre les mains pour sortir du cauchemar des relevés manuels, et il réalise ce à quoi pensait François RICHAUDEAU quand il imaginait un système avec plusieurs caméras-vidéos pour étudier les façons d'écrire.

Si l'avenir du module de *Lecture Méthodique* semble être celui d'un logiciel classique d'E.A.O., celui des deux autres modules est ouvert, les conclusions sont encore pour demain. La tâche est vaste et nous n'aurons pas trop de têtes pour y réfléchir. Ces travaux font l'objet de recherches INRP ainsi que de travaux de DHEPS pour certains d'entre nous. Les multiples voies ouvertes s'enrichissent mutuellement et n'enferment pas dans une discipline particulière. Elles permettent de travailler sur la construction d'un texte, en ouvrant de façon innovante sur les différents champs de la linguistique. ●

Denis FOUCAMBERT