

RAPPORT D'ACTIVITÉ DU CENTRE DE RECHERCHE POUR UN TRÉSOR DE LA LANGUE FRANÇAISE (1974-1976)

GÉRARD GORCY

I. POURSUITE DES TRAVAUX EN COURS

A — *Publications.*

1) Durant la période considérée ont été publiés les tomes 3 (*angelot-badin*) et 4 (*badine-cageot*) du *Trésor de la langue française, Dictionnaire de la langue du 19^e et du 20^e siècle*. Le tome 5 (*cageot-constat*) est sous presse et sera publié en mars 1977.

On notera que les tomes 4 et 5 du dictionnaire comptent plus de 1200 pages d'articles de dictionnaire et que le nombre de mots enregistré dans chaque volume sous la forme d'entrées, de dérivés ou remarques traités sous les entrées atteint maintenant environ 6000; ainsi pourra être rempli le programme prévu pour une quinzaine de volumes. Des fascicules complémentaires permettront d'illustrer d'exemples nombreux les mots les plus importants, notamment les mots polysémiques employés aussi dans divers domaines spéciaux (p. ex. en philosophie).

2) Le *Bulletin analytique de linguistique française* (BALF) a été régulièrement publié; les résumés ont été remplacés par des mots-clefs, principalement tirés du Thesaurus élaboré au laboratoire (cf. infra) et facilitant l'introduction en ordinateur des informations. Depuis avril 1976, il est diffusé par la librairie Klincksieck-Paris;

3) Un supplément annuel de 64 p. à la revue *Le Français moderne* a été publié; le supplément de 1975 a été consacré à la publication du *Thésaurus de linguistique française* établi par le service de documentation du laboratoire.

B — *Dépouillements et publications diverses.*

1) Les dépouillements mécanographiques de textes littéraires (au sens large du terme, comprenant donc aussi des textes «philosophiques») du dix-huitième siècle sont maintenant achevés (environ 28 millions de mots); ceux du 17^{ème} siècle sont en cours, une partie étant perforée (environ 8 millions de mots); une autre partie est en voie de sélection; à cette fin, a été distribuée aux participants du colloque une liste de textes provisoirement retenus qui sera publiée dans le supplément du *Français moderne* d'avril 1977; le concours des spécialistes est vivement sollicité pour affiner et enrichir une première sélection des oeuvres et des éditions qu'a préparée le service de bibliothèque du laboratoire.

2) Des index d'oeuvres ont été établis pour plusieurs chercheurs français et étrangers; on notera en particulier des index de pièces de théâtre contemporain (Giraudoux et Anouilh) établis à la demande du professeur Charles Muller (Université de Strasbourg) sous la forme d'index partiels des «discours» de chaque personnage; dans la même période des recherches syntaxiques ont été entreprises pour une équipe de Suède sur le verbe *faire*; enfin, pour une équipe de recherche travaillant sur le moyen français (animée par MM. P. Imbs et R. Martin de l'Université de Metz) différents travaux de lexicologie et de syntaxe ont été entrepris sur Guillaume de Machaut et sur le *Petit Jehan de Saintré*. Un *Onomasticon* complet de G. de Machaut a pu être ainsi établi et publié par un chercheur du laboratoire, M. Musso.

3) Différents documents ont été établis à l'usage des rédacteurs du laboratoire, et qui peuvent être consultés sur place; on signale particulièrement une *nomenclature des nomenclatures* du 19^{ème} et du 20^{ème} siècles (liste alphabétique et liste inverse) qui permet de comparer la nomenclature du T.L.F. avec celle des principaux dictionnaires généraux ou encyclopédiques des deux siècles ainsi qu'avec celle d'une nomenclature cumulative de plus de 500 dictionnaires techniques de la même période.

II. PRÉPARATION DE TRAVAUX EN VUE DU FUTUR INSTITUT DE LINGUISTIQUE FRANÇAISE (I.L.F.)

Cet Institut, dont la création est projetée pour octobre 1977, c'est-à-dire au moment du départ pour la retraite administrative de M.P. Imbs, sera dirigé par M.B. Quemada; ayant son siège au T.L.F. à Nancy, noyau central (mais non centralisateur) du nouvel organisme (on a créé à cet effet le concept administratif et juridique du *laboratoire-réseau*), il sera à la fois une base de regroupement et une banque commune de données puisqu'il vise à relier différentes formations de recherche en raison de la complémentarité de leurs programmes. Différents travaux ont d'ores et déjà été entrepris au laboratoire pour préparer ce nouveau développement et cette nouvelle orientation du laboratoire, où l'achèvement du dictionnaire T.L.F. restera prioritaire.

A — Travaux d'automatisation documentaire.

Ils ont été conçus en vue de rendre opérationnelle une base de données:

1) *données bibliographiques*: il est envisagé de publier le thésaurus de linguistique française (cf. supra) en traduction anglaise et de le compléter par un *thesaurus linguistique technique*.

2) *données documentaires*: il est prévu de publier un *Dictionnaire de fréquences du vocabulaire technique* dépouillé au T.L.F.; pa-

rallèlement, depuis mai 1975, fonctionne au laboratoire avec le concours scientifique de MM. Imbs et Muller une action thématique programmée (ATP), placée sous la responsabilité de M. Gorcy et intitulée: «Levée des ambiguïtés dans la langue écrite en français moderne par une lemmatisation semi-automatique». On voit l'intérêt d'un tel travail pour l'édition d'index d'oeuvres ou d'auteurs à partir du fonds T.L.F. ou de toute autre documentation textuelle sur le français moderne qui pourrait entrer dans la future banque de données.

B — Exploitation automatique de la documentation d'exemples pour la rédaction du dictionnaire.

Des stages de formation, organisés avec le concours de la Direction du C.N.R.S., ont visé à une démystification de l'utilisation de l'ordinateur par les rédacteurs.

Parallèlement au changement d'ordinateur, qui est maintenant un Iris 80 C.I.I., une amélioration du système d'exploitation de la documentation d'exemples a été entreprise, qui à un fichier-répertoire initial des textes enregistrés sur rubans magnétiques substitue un fichier-répertoire de mots. Au moment de la rédaction d'un article, il devient possible au rédacteur d'obtenir très rapidement des informations complémentaires. On soulignera à cet égard la valeur heuristique d'un listage qui, pour une unité lexicale étudiée, présente un classement des cooccurrences selon qu'elles sont le premier ou le deuxième mot qui précède, le premier ou le deuxième mot qui suit.

Le but recherché est donc l'utilisation par les rédacteurs de procédures conversationnelles avec les moyens de l'informatique déjà existants; l'implantation projetée au laboratoire d'un terminal lourd (du type Mitra 125 C.I.I.) sera décisive à cet égard.