

Version de travail

paru dans :

1998, Léon Jacqueline, “Les débuts de la traduction automatique en France (1959-1968): à contretemps?”, *Modèles Linguistiques*, tome XIX, fascicule 2 :55-86.

Jacqueline Léon

UMR7597, Laboratoire d'Histoire des Théories Linguistique

Les débuts de la TA en France (1959-1968): à contretemps?

Introduction¹

Le mythe de la machine à traduire précède l'apparition des ordinateurs et les premières machines à traduire remontent aux années 30 notamment avec la machine du soviétique Troianskji. Dès ce moment là elle est liée à la demande sociale de traductions en masse. Les premières expériences de TA, qu'on peut définir comme la traduction à l'aide d'une calculatrice électronique d'un texte dans une langue source donnée dans un texte dans une langue cible, datent de 1948 en Grande-Bretagne et aux USA. L'URSS commence en 1954², et l'Italie en 1959. La France ne commence qu'en 1959-60 avec la création de l'ATALA (l'Association pour l'étude et le développement de la Traduction Automatique et de la Linguistique Appliquée) en 1959 et du CETA (Centre d'Études pour la Traduction Automatique), en décembre 1959, au sein de l'Institut Blaise Pascal, comprenant deux sections, l'une à Paris dirigé par Aimé Sestier, l'autre à Grenoble dirigé par Bernard Vauquois. Enfin un Groupe de TA est créé en mai 1960 à la faculté des lettres de Nancy dirigé par Bernard Pottier et Guy Bourquin.

Nous partirons de deux questions afin d'étudier la spécificité de la recherche française en TA. Premièrement pourquoi la France a-t-elle commencé si tard, au moment même de la parution du rapport Bar-Hillel qui, je le rappelle, critiquait très fortement la TA aux USA, sur le plan des résultats comme des méthodes.

¹ Je tiens à remercier André Lentin, Yves Gentilhomme, René Moreau, Guy Bourquin, Bernard Pottier, Antoine Culioli, Maurice Gross, Sophie Fisher et Claude Del Vigna pour les informations qu'ils m'ont communiquées ou les archives qu'ils ont bien voulu me confier. Ils ne peuvent bien sûr être tenus pour responsables des erreurs ou omissions qui figureraient dans cet article.

Ma reconnaissance va particulièrement à André Lentin qui a relu ce texte et qui m'a toujours aidée et encouragée dans ce projet d'histoire de la traduction en France.

Je remercie enfin Louis Cosnier et Michèle Sabourin, du Service des archives du CNRS, pour la gentillesse de leur accueil et pour la très grande efficacité de leur aide.

² Pour les premiers travaux soviétiques, voir Archambault S. et Léon J., 1997, “La langue intermédiaire dans la Traduction Automatique en URSS (1954-1960)”

Une deuxième question concerne une deuxième date clef, 1967, c'est-à-dire un an après le rapport de l'ALPAC qui a marqué l'arrêt des subventions aux expériences de TA et signé la disparition de la plupart des centres faisant de la TA. Or cette année-là, le CETA qui n'a jamais été aussi florissant, se voit octroyer un nouveau bâtiment inauguré en août 67 sur le campus de St Martin d'Hères, et organise le second colloque sur le Traitement Automatique des Langues, où il présente la première démonstration de traduction effective russe-français sur ordinateur; treize ans donc après la première démonstration sur ordinateur aux USA qui a eu lieu en janvier 1954 à New York, à l'initiative de l'équipe de l'Université de Georgetown dirigée par Léon Dostert.

A travers ces deux questions, pourquoi la France a-t-elle commencé si tard et pourquoi semble-t-elle non affectée par le rapport de l'ALPAC, nous allons considérer la spécificité française de la recherche en TA de 1959 à 1970 en insistant ici sur les aspects institutionnels³.

Nous allons d'abord examiner les causes du retard français à travers la situation de l'informatique et de la linguistique en France à la fin des années 50, et le rôle crucial de l'ATALA dans les débuts de la TA en France. Nous développerons peu l'histoire de la création de l'ATALA, celle-ci ayant déjà fait l'objet de publications antérieures (cf. Lentin 1992, Léon 1992) même si un travail en profondeur reste à faire. Dans cet article, nous nous concentrerons donc sur les premières années du CETA, à travers lesquelles il est possible d'appréhender la position du CNRS dans les années 60 à l'égard de la traduction automatique.

1. Le retard français

1.1. L'informatique⁴ en France à la fin des années 50

Dans les années 50, il y avait une nette prédominance du calcul analogique sur le calcul numérique qui en était encore à ses balbutiements. La France est très en retard. En 1954, il n'existait aucune machine numérique en France, alors qu'en Europe, il en existait deux en Grande-Bretagne et une en Allemagne.

Ce retard tient à un ensemble de raisons. Henri Boucher (1988) raconte que de 1945 à 1951, l'informatique n'était aux USA qu'une simple expérimentation dont la renommée n'avait pas atteint la France. Ainsi aucune allusion ne fut faite à ces recherches dans les cours de l'École Polytechnique, de l'ENSGM, ni de Sup'Aero (écoles formant les militaires). Dans ce contexte, Sestier, qui prendra la direction du CETAP, était un des rares, à l'armement, qui s'intéressait au calcul numérique.

³ Une étude sur les aspects proprement scientifiques est en cours.

⁴ On ne parlait évidemment pas d'*informatique* à l'époque, le terme n'étant apparu qu'en 1962 sous la plume de Ph. Dreyfus - Dictionnaire des mots nouveaux, P. Gilbert, 1972 - , formé à partir de *information* et *automatique*. Il ne fut réellement en usage qu'à la fin des années 60.

Les diverses compagnies accoutumées à travailler avec la Défense sur les problèmes d'électronique et de mécanique de précision ont toutes refusé de prendre un risque technologique en se lançant dans la construction d'une machine française. Se reconnaissant incompetentes tant en architecture qu'en composants, elles ont proposé de s'associer à des sociétés américaines par le biais de licences et la Défense, particulièrement soucieuse du secret de ses études, n'a pas poursuivi. La seule société qui ait accepté de mener des études technologiques sans recourir à une compétence étrangère a été IBM France dont les prix étaient par ailleurs inférieurs à toutes les propositions des sociétés françaises.

Alors que l'armée et l'industrie françaises ne manifestent qu'un intérêt frileux pour les calculatrices électroniques, et que l'université fait preuve d'une très grande hostilité a priori à l'égard des machines et du traitement de l'information, le CNRS est le seul, à travers l'Institut Blaise Pascal, à se lancer dans un projet de construction de machine électronique. Hélas ce projet n'aboutira pas. Il s'agit de la fameuse affaire Couffignal et de la 'non construction en France du premier ordinateur' étudiée par G. Ramunni (1989, 1991, 1992).

1.2. L'Institut Blaise Pascal⁵

Afin de faire face aux besoins des laboratoires de recherche français en matières de calcul et pour développer la construction de calculateurs électroniques français, le CNRS crée en 1946 l'Institut Blaise Pascal (IBP) sous la supervision de Joseph Perès, alors directeur adjoint des sciences du CNRS (de 1945 à 1951). L'IBP⁶ regroupe deux laboratoires existants: le laboratoire de calcul analogique, dirigé par Lucien Malavard, et le laboratoire de calcul mécanique, dirigé par Louis Couffignal.

En 1947 le CNRS confie à Couffignal le soin de construire la première calculatrice électronique française, mais à cause d'une erreur de conception, la machine n'est pas techniquement viable et le projet est abandonné en 1952. Après ces précieuses années perdues, le CNRS décide finalement, en 1955, d'acheter une machine anglaise, une Elliott 402 pour équiper l'Institut Blaise Pascal. La première machine française est donc anglaise.

A son arrivée comme directeur du CNRS en 1957 (il le restera jusqu'en 1962), Jean Coulomb mettra fin aux fonctions de Couffignal qui sera remplacé par René de Possel (1905-1974). C'est aussi en 1957 que sera créé le Centre de mécanographie de documentation archéologique qui deviendra en 1969 le centre d'analyse documentaire pour l'archéologie (CADA), sous la direction de Jean-Claude Gardin.

⁵ Pour l'histoire de l'Institut Blaise Pascal, voir Mounier-Kuhn (1994).

⁶ Sur la naissance de l'informatique universitaire en France et les débuts de l'Institut Blaise Pascal, voir Nolin (1998).

Sont créés par la suite à l'IBP, le CETA en décembre 1959, puis la Section d'Automatique Documentaire (SAD) le 20 décembre 1960 sous la direction de Jean-Claude Gardin⁷. La SAD comprend alors Maurice Coyaud, Monique Renaud et R.C. Cros.

En 1962, à la mort de Joseph Perès, Malavard quitte Blaise Pascal et fonde ce qui deviendra, en 1972, le LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur). R. de Possel devient alors directeur de l'ensemble de l'IPB.

Toutefois l'IBP ne bénéficie plus du soutien du CNRS, en particulier de son directeur Pierre Jacquinot (de 1963 à 1969), et de son directeur scientifique Georges Jobert nommé en 1966. Le CNRS met tous ses efforts du côté de l'IRIA (Institut de Recherche d'Informatique et d'Automatique) et l'Institut Blaise Pascal ferme donc ses portes en 1969⁸.

1.3. Les machines et les hommes

En 1959, il n'existe que huit machines en tout et pour tout, pour le CNRS et l'Université. L'année 1959 marque toutefois un tournant. On assiste alors à un bond décisif dans le développement de l'équipe de la recherche en calculatrices électroniques grâce à la décision, supportée par une politique gaullienne favorisant la recherche, de développer prioritairement trois grands centres informatiques à Paris, Grenoble, et Toulouse. C'est donc là où sont les machines que seront créés les centres de TA.

Selon le projet préparatoire au rapport de conjoncture 1960 du CNRS, le retard français en informatique ne tient pas seulement au manque de matériel. Il tient aussi à une pénurie générale de personnel spécialisé. Il est difficile d'intéresser et de former les ingénieurs au calcul numérique. Par ailleurs, il faut assurer la reconversion des calculateurs sur machines de bureau à la programmation. Dans le cas du CNRS, précise ce rapport, la situation est particulièrement grave: il est devenu impossible de recruter des personnes qualifiées auxquelles l'industrie privée offre des situations beaucoup plus avantageuses.

Ce désintérêt est lié au fait que le calcul électronique, considéré comme une technique subalterne, n'est pris au sérieux ni par les universitaires ni par les entreprises.

1.4. La TA en France et la cybernétique

⁷ Gardin dirigera la SAD jusqu'en 1966, date à laquelle il demande à Maurice Gross de lui succéder, afin de reprendre ses activités de recherche. La SAD disparaîtra en 1970 au profit LADL (Laboratoire d'Automatique Documentaire et Linguistique) créé par Maurice Gross qui deviendra alors laboratoire associé au CNRS.

⁸ Il renaîtra en 1988 sous la forme d'une fédération de quatre laboratoires, le LADL, le LAFORIA (Laboratoire Formes et Intelligence Artificielle), le LITP (Laboratoire d'Informatique Théorique) et le MASI (Laboratoire de Méthodologie et d'Architecture Informatique).

On peut invoquer un autre élément pour expliquer le retard français en matière de TA. On sait que c'est le *Mémorandum* de Weaver, rédigé en 1949, qui a suscité les premières expériences de TA aux USA et en GB. Or Weaver est aussi vice-président de la fondation Rockefeller qui a financé tous les colloques internationaux du CNRS (38 plus précisément), de 1945 à 1955. Un de ces colloques aurait pu traiter de la TA et susciter un intérêt pour ce domaine en France. Il s'agit d'un colloque organisé par Couffignal, toujours lui, à l'IBP en janvier 1951 intitulé 'Les machines à calculer et la pensée humaine'. Les discussions portent sur la cybernétique et sur les machines à calculer numériques (cf. Ramunni, 1992), mais malgré la présence de A.D. Booth, un des pionniers britanniques de la TA, celle-ci n'est pas abordée et ce domaine ne semble pas trouver d'écho en France.

1.5. La linguistique au CNRS en 1959

Enfin, si l'on regarde la situation de la linguistique en France, la linguistique traditionnelle est orientée essentiellement vers la philologie, l'enseignement des langues et les prescriptions normatives⁹. Dans le rapport de conjoncture du CNRS (1959-60), la linguistique est représentée par deux sections : "linguistique et philologie classique" dont fait partie le directeur adjoint aux lettres et sciences humaines, Michel Lejeune, et la section "linguistique et philologie non-classique" à laquelle appartiennent Emile Benveniste, Marcel Cohen et André Martinet. Celle-ci deviendra en 1960 la section 22 "linguistique générale, langues modernes et littérature comparée" dont feront partie Benveniste, Delavenay, Martinet et Jean Perrot.

Il faut préciser que les linguistes français ne manifestent pas d'intérêt spécifique pour les langages formels, qui, aux USA, sont développés parallèlement aux expériences de TA, souvent au sein même des centres de TA. Chomsky, par exemple, jeune fellow de Harvard, spécialisé dans les problèmes de méthodologie en linguistique, est embauché en 1955 par Victor Yngve au laboratoire d'électronique du MIT pour travailler sur la TA avec une bourse de la NSF. Même s'il n'a jamais réellement travaillé sur un projet de TA, il a bénéficié de cet environnement de recherche. C'est notamment là qu'il écrira son premier article en 1956.

Toutefois nous verrons que certains linguistes se regrouperont autour d'Emile Delavenay pour impulser la recherche en TA en France et seront impliqués dans son développement.

2. Emile Delavenay, l'UNESCO et l'ATALA

2.1. La préhistoire de la TA en France: le rôle d'Émile Delavenay et de l'UNESCO

⁹ Sur la situation de la linguistique en France au début des années 60, voir entre autres l'article de Chevalier et Encrevé (1984) et le numéro de *Langue Française* n°117 consacré à 'La linguistique comme discipline en France'.

A la fin des années 50, la direction du CNRS n'est pas sans avoir une idée, au moins sommaire, sur les expériences en TA, du moins sur celles qui sont effectuées aux USA. Il semble qu'elle en ait été informée en 1956, lorsque Gaston Dupouy, directeur du CNRS (de 1950 à 1957), est convié, par l'intermédiaire d'un militaire, le Général Brison, à une réunion internationale "sur les machines à traduction" organisée par Léon Dostert, responsable du groupe de TA de Georgetown University, le 20 octobre 1956 au MIT. Il en informe le linguiste Michel Lejeune¹⁰, directeur adjoint en lettres et sciences humaines (de 1955 à 1963), lequel ne peut pas s'y rendre.

Un an après, en octobre 1957, l'attention de Lejeune est de nouveau attirée sur la TA par Henri Laugier, professeur à la Faculté des sciences, qui lui transmet une lettre exposant l'intérêt de monter un groupe de TA en France. Cette lettre est signée d'Emile Delavenay, directeur du service des documents et des publications de l'UNESCO.

Né en 1905, angliciste de formation et ancien normalien¹¹, E. Delavenay, par ses fonctions jusqu'en 1950 de responsable des services de traduction et d'édition à l'ONU à New York, a été amené à s'intéresser de très près aux problèmes de traduction sur le plan international. Il n'est donc pas étonnant qu'il ait été l'instigateur de la TA en France.

Finalement, en novembre 1957, Lejeune et Delavenay se rencontrent. Entre-temps Lejeune a fait la connaissance d'Erwin Reifler, sinologue et responsable du groupe de TA de l'Université de Washington, lors du 8ème congrès de linguistique d'Oslo, et est au fait des problèmes de TA.

A l'issue de cette entrevue, Delavenay convoque en février et mars 1958 deux réunions sur "la machine à traduire" à l'Institut de Linguistique de la Sorbonne en collaboration avec André Martinet. A cette réunion présidée par Benveniste, à laquelle assistera M. Lejeune, participeront les mathématiciens Georges Guilbaud, Benoît Mandelbrot, Marcel-Paul Schützenberger, Clemens Heller de la VIème section de l'École Pratique des Hautes Études, et Louis Ziéglé.

Dans ses mémoires, Delavenay (1992), évoque la faible réceptivité des linguistes, et des universitaires en général, à l'idée de fabriquer une machine à traduire en France et ces réunions ne semblent pas avoir abouti sur un projet concret.

Parallèlement, autour de Delavenay, se constitue un groupe de travail sur la TA qui se tient au courant des travaux américains, britanniques et russes. Ce groupe prend le nom de 'groupe international d'études sur la traduction automatique' et se réunit régulièrement à l'UNESCO. Il sera à l'origine de l'ATALA.

¹⁰ Michel Lejeune, né en 1907, ancien élève de l'ENS, est un spécialiste de grec ancien, des langues celtiques et italiques, qui s'est beaucoup intéressé à l'histoire des systèmes d'écriture.

¹¹ Coulomb, Lejeune et Delavenay ont été, à la même époque, élèves à l'École Normale Supérieure d'Ulm (cf. Delavenay, 1992).

Bien que les activités de ce groupe ne soient pas directement liées à la fonction de Delavenay dans l'UNESCO, le rôle de l'UNESCO dans l'affaire n'est pas négligeable. Les problèmes de traductions scientifiques et techniques intéressaient de très près l'UNESCO qui a produit deux rapports sur la question. Le premier datant de 1949 (cf. Helmstrom 1951) est consacré aux dictionnaires terminologiques spécialisés; le second, datant de 1953, dresse un état des lieux des besoins croissants en traductions scientifiques et techniques dans le monde, faisant état du manque de formation des traducteurs et du coût trop grand des traductions.

Par ailleurs, la création de l'ATALA en septembre 1959 a bénéficié du succès du premier congrès de l'IFIP (International Federation for Information Processing) organisé par l'UNESCO à Paris en juin 1959 où sont présentées de nombreuses communications sur la TA.

Enfin, c'est grâce à Delavenay et Michael Corbe, tous deux de l'UNESCO qu'Aimé Sestier, qui deviendra directeur du CETAP, prend connaissance des recherches en TA à l'étranger.

Delavenay continuera à jouer un rôle important dans le développement de la TA en France. Il est en relation avec de nombreux groupes de TA anglais et américains: le groupe de Locke au MIT, le groupe de Reifler à l'Université de Washington, et le groupe de Cambridge en Grande Bretagne. Président fondateur de l'ATALA jusqu'en 1965, auteur d'un *Que sais-je?* intitulé *La machine à traduire*, et d'une bibliographie sur la TA (cf. Delavenay 1959 et 1960), il fera partie du Conseil scientifique du CETA et de la section 22 'Linguistique générale, langues modernes et littérature comparée' du CNRS.

2.2. L'ATALA¹²

Le bureau fondateur de l'ATALA, présidé par Delavenay, comprend l'Ingénieur Général Pierre Nicolau, directeur de l'Institut Supérieur des Matériaux et de la Construction mécanique, et directeur de la revue *Automatisme*, Pierre Meile, sanskritiste et professeur de langues modernes de l'Inde à L'École Nationale des Langues Orientales vivantes, Georges Fargues angliciste et spécialiste de l'Afrique, André Lentin, né en 1913, alors professeur de mathématiques supérieures, gendre de Marcel Cohen et bon lecteur du russe, Iréné Poncet, ancien professeur de physique et colonel dans l'armée, Claude Métais, professeur d'anglais intéressé par l'enseignement des langues.

¹² Sur la création de l'ATALA et sur la biographie de ses membres fondateurs, voir Lentin 1992.

Parmi les membres fondateurs, il faut compter également le linguiste Georges Gougenheim, père du 'français élémentaire', le linguiste Alexis Rygaloff sinologue et russophone, S. de Piccioto, ingénieur mis à la disposition de l'UNESCO par IBM auquel succéda René Tabory, Jacqueline Poyen ingénieur de la compagnie Bull, Michael Corbe, journaliste international et traducteur à l'Unesco, René Moreau, directeur scientifique d'IBM, Paul Braffort mathématicien et futur fondateur d'Oulipo (novembre 1960), Jean Esmein, ancien lieutenant de vaisseau, ayant étudié le chinois et le japonais à l'École des Langues Orientales et membre de la compagnie Bull.

Font aussi partie des membres fondateurs de l'ATALA, Bernard Vauquois, Jean-Claude Gardin, et Yves Gentilhomme, tous pionniers de la TA et de la documentation automatique sur lesquels nous reviendrons plus loin dans cet article.

Outre des traducteurs et des mathématiciens, l'ATALA comprend de nombreux linguistes parmi ses membres: outre G.Gougenheim et A. Rygaloff déjà cités, Marcel Cohen, Jean Fourquet, Antoine Culioli, Gilbert Lazard, Bernard Pottier et David Cohen entre autres. Fin 1960, elle comprendra une centaine de membres.

L'ATALA se considère non comme un organisme de recherches mais plutôt comme un forum de discussion sur la TA mais aussi sur la documentation automatique et la linguistique appliquée. Ce dont témoigne sa revue créée en avril 1960 *la Traduction Automatique* (infra *la TA*), fabriquée alternativement par les Compagnies Bull et IBM impliquées dans les projets de traduction automatique.

La TA se veut moins une revue scientifique (1/3 d'articles et 2/3 d'informations, comptes-rendus et bibliographies¹³) qu'un bulletin de liaison, d'informations et de formation. Elle rend compte de façon systématique de toute manifestation et de toute publication dans le domaine du traitement non numérique de l'information C'est un des lieux de pénétration en France de la linguistique américaine et des langages formels grâce notamment aux comptes-rendus systématiques des ouvrages dans *la TA* dès leur parution.

Signalons également en Mars 1960, la création du Séminaire de Linguistique Quantitative dirigé par Jean Favard, à l'Institut Henri Poincaré. Y est diffusé un enseignement de linguistique pour mathématiciens (André Martinet, Jean Dubois) et un enseignement, pour non-spécialistes, de mathématiques, logique mathématique et théorie de l'information (René Moreau et Daniel Hérault).

En 1962-1963, un cours supplémentaire sur la théorie des langages sera assuré par J. Pitrat et M. Gross. A l'origine de ce séminaire, un groupe fondé par le Ministère de la Défense après l'aventure de Suez pour étudier le codage des messages à l'aide de modèles statistiques; certains de ses membres, comme René Moreau, alors Capitaine de

¹³ *La TA* emprunte à son homologue américaine the *Journal of Mechanical Translation*, fondée en 1954, sa présentation bibliographique sous forme de résumé et la transforme en fiches détachables.

Gendarmerie, ayant travaillé pour le chiffre pendant la guerre et qui deviendra Directeur du centre scientifique d'IBM.

3. Création du CETA

En 1959, l'intérêt du CNRS pour les recherches en TA se confirme. Dans le rapport de conjoncture de 1959 du CNRS, la TA figure parmi les sujets prioritaires que se donne la commission de mathématiques générales et appliquées, présidée par le Doyen Péres, en collaboration avec la commission de physique théorique et de théorie des probabilités. Celle-ci fait état d'une demande en traduction automatique du russe, en documentation automatique, en analyse numérique et en automatisme.

A cette fin il faut équiper en calculateurs électroniques deux ou trois super-centres: Paris, Grenoble et Toulouse, avec des Gamma 60 ou des IBM 704. C'est à ces centres que seront confiés l'étude des problèmes prioritaires.

Par ailleurs, la Direction de l'Armement, alertée par le Général Grossin, alors directeur du SDEC (Service de documentation et de contre-espionnage), commence à s'intéresser à la traduction automatique et se montre disposée à y consacrer une partie de son budget de recherche.

Le CETA est créé le 24 décembre 1959 par une convention entre le CNRS, la DEFA *Direction des Etudes et Fabrications d'Armement* et le CASDEN *Comité d'Action Scientifique de Défense du Centre d'Exploitation Scientifique et Technique* du Ministère de la Défense¹⁴. Jean Coulomb, directeur du CNRS de 1957 à 1962, représente le CNRS, le Général Guérin représente le CASDN et le Général Sorlet la DEFA. Il est créé, sous la supervision du Doyen Péres, au sein du Laboratoire de calcul numérique de l'Institut Blaise Pascal du CNRS. Il a pour mission "l'étude et la conception d'une méthode pour la traduction automatique notamment du russe en français et l'étude de l'organisation générale d'une machine pour cette fin".

La création du CETA à l'Institut Blaise Pascal associe donc étroitement la TA au développement du calcul numérique et des calculateurs électroniques ainsi qu'à la documentation automatique. Toutes deux sont sous le contrôle des mathématiques appliquées et dépendent de la Section 'Mécanique générale et mathématiques appliquées' du CNRS. Même si les équipes sont composées de spécialistes des diverses disciplines intéressées, les mathématiciens appliqués considèrent qu'ils sont les seuls à pouvoir animer la recherche, dans la mesure où ils les seuls capables d'assurer le suivi des questions de programmation et de structure des machines.

Le directeur du laboratoire de calcul numérique, René de Possel, et son sous-directeur, André Lentin, joueront un rôle important dans l'interaction entre mathématiques

¹⁴ Le CASDEN était déjà en charge de la traduction d'articles scientifiques russes.

appliquées, langages formels et linguistique. Ils sont tous deux membres du Conseil scientifique du CETA.

3.1. Le CETAP

L'armée est, à plusieurs titres, tout à fait partie prenante dans l'affaire. Outre l'aide du CASDN, qui assure pour moitié le financement de l'ensemble du CETA, le CETAP bénéficie des locaux du Laboratoire Central de l'Armement (LCA), au Fort de Montrouge à Arcueil. Son directeur, Aimé Sestier, né en 1920, ancien élève de l'École Polytechnique, est ingénieur militaire en chef, Chef de la Section des machines à calculer du LCA. L'armée est très présente dans le comité de direction du CETA: outre des représentants du CASDEN comme le général Guérin, et du LCA comme le général Sorlet, le comité de direction comprend l'ingénieur général Ollier représentant le SDEC¹⁵.

Les ingénieurs et le personnel technico-administratif appartiennent au LCA. Les deux ingénieurs à plein temps du CETAP sont Lucien Dupuis, ingénieur civil à la DEFA, qui par ailleurs connaît le russe, et Maurice Gross, né en 1934, ancien élève de l'École Polytechnique et ingénieur de l'École Nationale Supérieure de l'Armement. Celui-ci est attaché au service des machines à calculer du LCA. Des ingénieurs, programmeurs et apprentis programmeurs se joindront à eux, la plupart du temps à temps partiel, à partir de février 1960: M. Baloffet, Mr Etienne, Nguyen Can, M. Berlicot, L.Hennebutte, Bernard Jaulin, C. Precetti.

Les linguistes sont rémunérés par le CNRS. Cependant, à l'exception d'Yves Gentilhomme, les russisants engagés par le CETAP ne sont pas des chercheurs, ce sont des agrégés de russe et parfois même de simples locuteurs natifs du russe. On dispose sur ce point le témoignage d'une des linguistes russes engagée au CETAP sous la direction d'Y.Gentilhomme: arrivée en France à 18 ans en 1957 avec un "baccalauréat" russe qui comprenait une légère formation paramilitaire, elle est affectée à la traduction de listes de termes militaires russe-français.

Yves Gentilhomme, né en 1920, de langue maternelle russe, est détaché de son poste de professeur de mathématiques au Lycée de St Germain en Laye comme chercheur au CNRS à partir de 1961. Six linguistes slavisants travaillent sous sa direction: Olivier Tailleur, Melle Petrenko, Melle Baconnier, Catherine Paris, Melle Toubnikoff, Mme Vichniakoff. Enfin on compte deux linguistes germanistes: Jacques Perriault, né en 1938, licencié en droit et en sciences économiques et diplômé de programmation chez IBM, et Madame Dupuis.

¹⁵ Le Comité de direction, présidé par Jean Coulomb, directeur du CNRS, comprenait également le Doyen Pérès, E. Delavenay, R. de Possel, M. Lejeune, Raymond Cheradame, directeur des études de l'École Polytechnique, A. Sestier et B. Vauquois.

3.2. Le CETAG

Grenoble, on l'a dit, est un des principaux lieux d'implantation de l'informatique en France. C'est à Grenoble qu'est fondée l'AFCAL (Association Française de Calcul) en 1957 et Kuntzmann y crée la revue 'Chiffres' en 1958.

Il n'est donc pas étonnant que la seconde section du CETA soit créée à Grenoble. Et c'est au laboratoire de calcul de la faculté des sciences, avec le soutien du professeur Kuntzmann qu'est créé le CETA sous la direction de Bernard Vauquois.

Bernard Vauquois (1929-1985) est d'abord chercheur au CNRS de 1952 à 1958 en radio-astronomie au Laboratoire d'astronomie à l'institut d'astrophysique de l'observatoire de Meudon, après avoir effectué des études de mathématiques, de physique et d'astronomie. Dès 1957, son programme de recherche porte aussi sur les méthodes appliquées à la physique du point de vue des calculateurs électroniques et il enseigne la programmation aux physiciens. Ce double intérêt pour l'astrophysique et les calculateurs électroniques se reflète dans ses deux sujets de thèse d'état en sciences physiques qu'il soutient en 1958: (i) étude de la composante lentement variable du rayonnement radioélectrique solaire, (ii) étude de la théorie des machines (Machine de Turing et Von Neumann). Il est nommé maître de conférences en mathématiques appliquées à la faculté de sciences de Grenoble en octobre 1958.

Le groupe de réflexion sur la TA qui deviendra le CETAG est composé de mathématiciens et d'ingénieurs mais manque cruellement de linguistes, notamment de russisants. Le groupe bénéficie toutefois du conseil d'un certain nombre d'agrégés de langue. Selon Vauquois ce sont les linguistes qui ont adhéré au projet des mathématiciens et ingénieurs du Laboratoire de calcul. Vauquois, dans le rapport pour le conseil scientifique du 20/02/60, souligne le gros atout du groupe: l'accès à la machine Gamma ET à tambour magnétique.

Après une demande spécifique de linguistes effectuée par Vauquois en mai 1960 auprès de la direction du CNRS, le CETAG comprend, à la fin de la même année, quatre spécialistes de russe: un attaché de recherche, Nicolas Pogarieloff, agrégé de russe, et trois techniciennes: Dominique Andron, Lydie Torre et Anita Gagny. Il comprend une technicienne germaniste, Gisela Laurent, et une technicienne spécialiste de japonais Yamada Sae. En ce qui concerne les informaticiens, le directeur de la programmation, l'ingénieur Jean Veyrunes, disparu prématurément en 1966, a sous ses ordres deux programmeurs: Ernest Grandjean et Marcel Brissaud, et une opératrice. Le groupe compte également deux ingénieurs mathématiciens préparant un doctorat de mathématiques appliquées, Alain Auroux et Gérard Veillon.

4. Le rapport Sestier

Bien que recevant des crédits équivalents, les deux groupes ne sont pas à égalité au départ. Alors que le groupe qui constituera le CETAG vient de commencer les recherches en TA (octobre 1959), le CETAP dit avoir travaillé sur l'allemand depuis 1958, en tant que groupe du LCA. Mais surtout, Sestier a déjà rédigé un rapport circonstancié sur sa conception de la TA intitulé *Comment doit être organisé à l'échelle française l'effort pour la TA* (rapport Sestier du 23-11-59) C'est ce rapport qui va servir de base à la définition des orientations du centre.

4.1. Un centre de traduction en séries

Selon ce rapport, le CETA doit poursuivre un objectif pratique. D'ici un délai d'environ cinq ans, le CETA a pour vocation d'être à la fois une usine de production en série de versions françaises de russe scientifique et technique, un centre de recherches appliquées où des équipes constituées prépareront la TA d'autres langues, et un centre de recherches fondamentales en linguistique sans objectif immédiat de TA avec l'espoir que les résultats de la TA, complétés par des études indépendantes pourront être utiles à des linguistes pour élaborer ou contrôler des théories.

Pour Sestier, la production en série de traductions est donc au premier plan et le CETA devra offrir un certain nombre de prestations: traductions grossières et études sur l'indexation et l'extraction automatique susceptibles d'assurer à terme, d'ici cinq à dix ans, l'autofinancement des recherches¹⁶. Sestier propose d'ailleurs le nom de *Centre d'études et d'exécution de traductions automatiques* qui souligne sa vocation de réponse à la demande sociale.

Sestier conclut son rapport par une mise en garde concernant le risque de gaspiller les efforts et les crédits en ne faisant que redécouvrir ce qui a déjà été trouvé aux USA et en URSS. Sestier, ni personne, ne semble alors être au courant de la mission d'évaluation des centres de TA confiée en 1958 par la National Science Foundation américaine à Bar-Hillel, et dont les résultats seront publiés un peu plus tard, en 1960, connus sous le nom de 'rapport Bar-Hillel'.

Il faut noter que le rapport de conjoncture du CNRS (1959-60) reprend à son compte l'idée que dans un délai de cinq à dix ans, on ait dépassé le stade de la recherche pour produire des traductions en série.

4.2. Technologie linguistique et ingénieurs linguistes

La méthode préconisée par le rapport Sestier est surtout axée sur l'analyse de la langue source. Il s'agit du russe bien sûr, comme pour la plupart des travaux américains, et la

¹⁶ R. de Possel va s'attaquer à un des problèmes qui risque selon lui de freiner cet autofinancement, à savoir la lecture automatique des caractères. Cette machine à lire restera un centre d'intérêt majeur(cf. Journée d'hommage à de Possel, 1994).

tâche à accomplir est la traduction d'articles scientifiques et techniques. Les objectifs sont en effet en partie liés à l'armée et au contre-espionnage¹⁷.

Une lettre de Sestier du 18/10/60 à de Eric de Grolier, directeur du Centre français d'échange et de documentation technique à Milan, expose la méthode qu'il préconise: la TA peut se passer de théorie parfaite, ce qu'il lui faut c'est une 'technologie linguistique'. Comme Delavenay (1960b), Sestier parle de machine à traduire française. Cette technologie linguistique doit être le fondement scientifique de toute TA raisonnablement correcte, même si elle ne vise pas l'idéal bien lointain de la FAHQMT (Full Automatic High Quality Machine Translation) préconisée au début des années 50 par certains chercheurs américains. Les recherches doivent s'attacher dans un premier temps, moins à construire la morphologie, la syntaxe ou la sémantique de telle ou telle langue particulière, qu'à dégager la structure que doivent avoir ces disciplines, en tant que destinées à servir de base à la TA; et cela par des amorces d'études comparatives de diverses langues.

Dans son rapport, Sestier reprend les propositions qu'il a faites dans son article "*La traduction automatique des textes écrits scientifiques et techniques d'un langage dans un autre*" publié en 1959.

Les outils habituels de la traduction manuelle, dictionnaires, grammaires et syntaxes sous leur forme usuelle sont inadéquats. La traduction entre langues naturelles doit s'inspirer de la traduction entre langages artificiels et des méthodes de programmation automatique, où le langage de départ est un système de notation algébrique et le langage d'arrivée est le code d'instructions du calculateur automatique. Même si la traduction entre deux langues naturelles est beaucoup plus complexe, une grammaire formelle de la langue source est nécessaire.

Aucun procédé n'a permis de construire une grammaire purement formelle raisonnablement complète pour aucune langue naturelle. Or on ne saurait attendre, pour commencer les recherches et les réalisations de TA, qu'il existe une telle grammaire pour la langue de départ choisie. Force est de rechercher des procédés plus rapides pour construire les grammaires très particulières nécessaires à la TA.

C'est l'objectif de la technologie linguistique. Celle-ci consiste à inventorier et à classer des faits linguistiques, syntaxiques, morphologiques et lexicaux, à partir d'un corpus de

¹⁷ Outre les enjeux purement politiques et militaires, renforcés dans un contexte de guerre froide, les Américains étaient persuadés que les Soviétiques étaient très en avance sur eux sur le plan scientifique, non seulement dans le domaine de l'aérospatiale (le premier spoutnik date de 1957) mais aussi dans le domaine de la TA. Cette opinion était partagée par les Français si l'on en juge par la première version d'une demande de subvention adressée par Sestier à l'OTAN en juin 1961 qui se termine par le constat que, bien que l'effort des Soviétiques en matière de TA soit, de l'avis même des experts américains, encore plus grand que celui des USA, on n'a, depuis 1958, que peu de renseignements sur les travaux russes. Une des hypothèses est que "les soviétiques préparent à l'Occident une surprise spectaculaire, telle, peut-être que l'édition de traductions-machine de livres soviétiques" (rapport cité p.28).

textes à traduire. Les classifications doivent tenir compte de plusieurs critères; ainsi, la classification des verbes doit se faire selon des traits communs à tous les verbes (par exemple la conjugaison) et aussi selon des traits particuliers à chaque classe (par exemple la nature et l'organisation du contexte sont différentes pour les diverses classes de verbes transitifs et pour les diverses classes de verbes intransitifs).

L'exécution des tâches matérielles exigées par ce travail de classification ne peut être menée à bien dans des délais raisonnables qu'avec l'aide de calculateurs automatiques: des linguistes analysent des fragments de textes sans trop s'occuper de les mettre en ordre; les informations résultant de leur travail sont introduites dans un ordinateur qui fournit des éditions de tous les catalogues possible consignant les résultats.

Le problème qui incombe principalement au linguiste-analyste est de définir les critères de classification sous forme de catégorie en évitant de définir ces catégories comme exclusivement formelles, sous prétexte que seule la forme est accessible à un ordinateur. Il faut aussi définir des catégories de la signification.

Cette méthode de technologie linguistique demande une organisation particulière du travail qui suppose trois groupes d'intervenants: les théoriciens (les linguistes savants), les expérimentateurs (les linguistes exécutants) et les consultants (les traducteurs scientifiques et techniques).

Cette ingénierie linguistique avant l'heure trouvera un écho dans les rapports des sections de linguistique du CNRS (rapport de conjoncture CNRS 63-64) qui déclarent que la TA, comme la documentation automatique, arrivées au stade de l'exploitation rentable, doivent faire appel à des 'ingénieurs-linguistes' comme on dit des 'ingénieurs-chimistes'. Le rapport signale toutefois que le problème principal, pour le développement de la linguistique en France, reste la pénurie de collaborateurs pourvus d'une formation linguistique 'fondamentale'

4.3. La répartition des tâches: un problème de machines?

Le rapport de Sestier prévoit une répartition des tâches entre les deux équipes, parisienne et grenobloise, répartition fondée officiellement sur les différences de matériel et les recherches entreprises avant le 1er janvier 1960. Selon cette répartition, Grenoble est chargée de la morphologie considérée comme plus facile et plus rapide à réaliser, et Paris, dont l'expérience est plus grande, se charge de la syntaxe. Puisque leur tâche est "simple", l'équipe de Grenoble décide de s'attaquer aussi aux problèmes de polysémie lexicale.

Décision décrétée provisoire et fluctuante dès la première réunion du Conseil scientifique du 20 février 1960. Martinet et Benveniste, qui en font partie, critiquent vivement cette division entre morphologie et syntaxe. Celle-ci n'est pas pertinente, disent-ils, dès lors qu'il s'agit de comparer deux structures: il y aurait intérêt à partir

d'une solution moins graphique et plus linguistique, c'est-à-dire de l'autonomie du syntagme, nominal et verbal.

Cette belle répartition va de plus se trouver très rapidement contrariée. Les prétextes en sont les différences de machines autant que les différences de méthode.

Selon Sestier (juillet 1960), le groupe de Grenoble développe un système morphologique adaptable uniquement sur machine binaire, donc strictement inutilisable pour le CETAP qui dispose d'une machine décimale de taille mémoire réduite. Par ailleurs le système du CETAG lui paraît inutilement compliqué. Les Parisiens décident donc d'affronter directement l'analyse morphologique. Il insiste toutefois sur l'harmonisation indispensable des deux systèmes d'analyse morphologique développés indépendamment par les deux équipes.

Pour leur part, dans leur projet 12 décembre 1960, les membres du CETAG montrent leur intention de faire aussi des recherches en syntaxe sur la traduction russe-français, japonais-français et allemand-français, en adoptant principalement le modèle que Sydney Lamb a développé dans le groupe de l'Université de Berkeley.

Pour revenir aux calculateurs utilisés, il y a un déséquilibre certain entre les équipes. Pour traiter des programmes de TA, il faut des machines à grande capacité mémoire. L'avantage est cette fois-ci au CETAG qui ne dispose au début que d'une machine Bull, binaire-décimale à mémoire-tambour de grande capacité, mais qui pourra dès la fin 1960 utiliser le GAMMA 60 et l'IBM 7090 du Centre de Calcul de Grenoble, qui comptent parmi les machines les plus puissantes de l'époque.

Le CETAP ne dispose que d'une 650 IBM, machine décimale à 2000 mots avec bandes magnétiques 727. Sestier adresse différents rapports au CNRS sur l'insuffisance des machines. Plusieurs solutions sont envisagées. Le Comité de direction du 17 décembre 1960, décide que soient examinées les possibilités offertes par la machine MarkII mise au point par l'US Air Force en collaboration avec le groupe de TA de Washington; une autre solution consiste à envisager la construction d'une machine à traduire spécifiquement française en donnant des directives aux constructeurs IBM et Bull.

Dans un rapport de novembre 1960, Sestier préconise l'utilisation de MarkII, machine spécifiquement dédiée à la traduction automatique, intégrant une mémoire photoscopique de grande capacité à un ordinateur IBM classique. Vauquois, dans un rapport de janvier 1961 se montre moins enthousiaste à l'égard de la machine Mark II qui, certes, présente l'avantage d'une consultation rapide de dictionnaires, mais qui est très peu adaptée à l'analyse syntaxique et sémantique, dans la mesure où elle ne peut dépasser le stade de l'étude de mots consécutifs qu'au prix d'énormes difficultés alors que le ordinateur classique offre des possibilités immédiates.

Il ne semble pas que le projet d'utilisation de MarkII ait abouti, malgré la visite qu'ont effectuée Sestier et Vauquois au centre de l'US Air Force à Rome dans l'Etat de New

York, lors de leur mission aux USA de mai-juin 1961. Même si la solution adoptée provisoirement a été de renforcer les machines existantes (la 650 IBM a été complétée en mai 1961 par des disques magnétiques 355 à très grandes capacité mais d'accès relativement lent), l'idée de construire une machine à traduire française n'est pas abandonnée (cf. §7.1 infra).

5. La dissolution du CETAP

L'abandon du travail en linguistique fondamentale, puis de l'idée qu'il suffit d'une technologie linguistique pour réaliser des dispositifs fonctionnels de TA, conduisent Sestier à démissionner et à dissoudre le CETAP en octobre 1962.

5.1. L'abandon de la recherche en syntaxe

La position de Sestier se radicalise¹⁸, lorsqu'en automne 1960, le CETAP décide d'arrêter les recherches sur la syntaxe 'à la main', en raison de son rendement nécessairement faible, pour se consacrer à la mise au point d'outillage informatique pour linguistes. Sestier désigne ainsi la méthode traditionnelle des linguistes, qui consiste à produire des analyses, sans aucune aide automatique, de phrases spécialement sélectionnées pour faire apparaître des difficultés nouvelles et pour tester la validité des modèles mathématiques de la syntaxe des langues européennes (allemand, français et anglais).

Or "le linguiste chargé d'une analyse doit sans cesse disperser son attention entre des structures considérées comme déjà élucidées et cataloguées, et des structures nouvelles. Il doit souvent, en face de ces dernières, remettre en question les précédentes. Un seul linguiste peut péniblement maintenir dans le temps son accord avec lui-même, pour éviter que ses propres archives soient vite périmées par contradictions internes. Pour une équipe entière cela devient pratiquement impossible." (Sestier, 1961b, p.4).

De plus, tous les groupes étrangers travaillant sur la syntaxe, dit-il, se heurtent aux mêmes difficultés théoriques, quant à la validité des modèles utilisés. Même la construction directe d'un modèle plus général, fondé sur une structure en réseau, non arborescente¹⁹ qui semble la plus adaptée, suscite d'insolubles discussions linguistiques.

¹⁸ Elle est exposée dans sa communication au deuxième congrès de l'AFCALTI en octobre 1961 (Sestier, 1961c). Il y reproche son indiscipline à Gross qui voulait faire de l'analyse syntaxique automatique malgré les avertissements de Sestier.

¹⁹ Sestier fait sans doute allusion ici aux travaux du groupe de Cambridge (UK), qu'il cite par ailleurs (rapport OTAN, Sestier 1961a) dirigé par Margaret Masterman qui a commencé les recherches en TA en 1954-1955. Ce groupe a développé un modèle sémantique de TA fondé sur la théorie des treillis (cf. Léon, 1998, à paraître). Ce modèle repose donc une structure non arborescente en réseau, comme celle préconisée par Sestier.

Un peu plus tard, dans son article "Programme des conflits, modèles des conflits" (*La TA* d'octobre 1960), Y.Lecerf compare à des fins d'analyse syntaxique automatique les arbres de Tesnière et ceux de Chomsky.

Sestier discute essentiellement les modèles de Chomsky et celui de Hays, Le modèle de Chomsky est une structure arborescente dont les noeuds sont les syntagmes et qui n'admet que des constituants continus. Quant au 'dependency model' de Hays, il repose sur une structure arborescente dont les noeuds sont exclusivement des mots isolés. Par ailleurs il se montre très intéressé par le modèle développé par Silvio Ceccato à Milan.

Sestier poursuit: "A la réflexion, rien de bien étonnant à cela: la linguistique est une science aussi ancienne que les mathématiques (4000 ans au moins). Si les spécialistes les plus chevronnés continuent à discuter de ses principes mêmes, c'est sans doute qu'en quelques mois, on ne peut, avec une équipe de jeunes chercheurs tout juste sortie de sa phase d'organisation, faire plus que des générations de savants; même en entreprenant ces recherches avec des idées de mathématicien ou d'ingénieur: bref en essayant d'examiner les faits linguistiques sous un angle neuf. C'est ce que prévoyaient, dès le départ, les maîtres de la linguistique française qui nous ont encouragés et conseillés, sans quelque secret amusement de notre folle entreprise." (opus cité p.5)

Il décide donc en automne 1960 de se consacrer à la morphologie du russe et à la mise au point d'outillages de recherche, d'aides automatiques pour les chercheurs linguistes, sous forme d'un système d'analyse automatique approximatif, susceptible de retouches successives.

5.2. Les linguistes français qui s'intéressent à la TA

Qui sont ces maîtres de la linguistique française mentionnés par Sestier? Les linguistes français sont présents dans les instances d'évaluation du CETA mais ne participent pas directement aux recherches. Outre Martinet, directeur de recherche d'Yves Gentilhomme, et Benveniste, le Conseil scientifique du CETA comprenait en 1960 Michel Lejeune, Marcel Cohen, Georges Gougenheim, Bernard Quemada et deux slavistes, Jean Train et Marc Vey. Plus tard Jean Fourquet en fera partie.

Alors que la TA et l'étude des langages formels sont tenues à l'écart du cercle de la linguistique officielle, la SLP (Société de Linguistique de Paris), où l'on se méfie de cette "linguistique pour ingénieurs", un certain nombre de linguistes s'intéressaient vraiment à la TA et faisaient partie de l'ATALA comme Georges Gougenheim et Marcel Cohen qui y étaient particulièrement actifs. Martinet, Culioli, Pottier, David Cohen et Georges Mounin sont aussi très tôt membres de l'ATALA. La TA permettait, selon eux, de poser de véritables questions linguistiques. Ainsi l'ATALA organisera un colloque sur le mot en décembre 1962, au cours duquel un certain nombre de problèmes suscités par l'utilisation de la machine sont évoqués: notamment la question de savoir si l'on peut donner une définition du mot en linguistique générale ou seulement dans le cadre d'une langue donnée.

Au CNRS, la TA est mentionnée, dans les années 63-64, comme appartenant à la linguistique appliquée et encouragée à ce titre.

Il serait donc faux de dire que la TA en France n'a suscité aucun intérêt de la part des linguistes. Mais cet intérêt, bien que manifesté par les plus novateurs, reste distant et sans implications dans des projets de recherche effectifs.

5.3. Le CETA dans le monde au début des années 60.

Avant la démission de Sestier, les deux groupes ont pourtant tenté de donner une envergure internationale à leurs recherches. Ainsi Sestier et Vauquois entreprennent en mai-juin 1961 une mission de cinq semaines aux Etats-Unis où ils visitent les principaux centres de TA, dont, on l'a vu précédemment, la machine Mark II de l'US Air Force. Les relations sont particulièrement bonnes avec le groupe de Reifler. Cette mission est suivie d'une réunion en septembre 1961, regroupant, dans les locaux du CETAP, un certain nombre de responsables de centres de TA: Victor Yngve du MIT, Paul Garvin de Ramo-Wooldridge, Kenneth Harper et David Hays de la Rand Corporation, D.F Meyers de National Bureau of Standards, Richard See de la National Foundation of Sciences et un représentant du groupe de Silvio Ceccato à Milan, M.Maretti.

Pendant cette période, les deux centres produisent régulièrement des documents sur leurs recherches: pour le CETAP, l'étude de la morphologie et de la syntaxe du russe et de l'allemand, ainsi que des algorithmes de consultation de dictionnaires; pour le CETAG l'étude de la morphologie du russe, de l'allemand et du japonais.

A partir de 1963, le CETAG étendra son champ d'intérêt à la morphologie du français, la constitution de dictionnaires, et à l'analyse syntaxique du russe, du japonais, de l'allemand et du français. Par ailleurs il s'intéressera à la "théorie des catalogues", élaborée au départ par Martin Kay comme système d'organisation d'une base de données. Le groupe de B.Vauquois partage avec celui d'A.Sestier une conception de la traduction entre langues naturelles assimilée à celle des langages formels et de programmation. Dans les années suivantes, Bernard Vauquois développa dans la mise au point de systèmes de TA l'analogie entre traduction et compilation. Très tôt en effet, grâce à l'intérêt particulier de B. Vauquois pour les langages formels, le CETAG donna une importance à la formalisation dans le développement des systèmes de TA. Cet intérêt fut renforcé lors de la mission de B. Vauquois et J. Veyrunes en Union Soviétique, en novembre-décembre 1962, auprès des groupes de TA. Lors de cette mission, les Grenoblois prennent contact avec Igor Mel'cuk et le groupe d'Andreev qui travaillent sur une langue intermédiaire (ou langue-pivot, comme la nomme Vauquois dans son rapport de mission). Ce voyage en URSS sera suivi de missions en août, septembre et octobre 1963 en Argentine, au Mexique et aux USA. Ces dernières

missions indiquent la vivacité toujours constante du groupe de Grenoble malgré la démission de Sestier²⁰.

Par ailleurs les deux groupes français participent aux différents congrès qui se tiennent à l'époque dans le domaine. Un congrès à Grenoble, en septembre 1960, où Dupuis, Sestier présentent des communications; le colloque de Teddington *Traduction automatique et linguistique appliquée*, en septembre 1961²¹; l'AFCALTI à Paris en octobre 1961; l'école d'été organisée par l'OTAN *Automatic Translation of Languages* en juillet 1962; enfin l'International Federation for Information Processing (IFIP) qui s'est tenu à Munich en août 1962.

Malgré cette présence du CETA sur la scène internationale, Sestier va démissionner et abandonner la TA.

5.4. La démission de Sestier

Plusieurs événements accélèrent le processus de démission de Sestier. L'année 1962 correspond au moment où le financement du CETA par l'armée se trouve compromis par le remplacement du CASDEN par le CASD, dépourvu de ressources propres. De Gaulle avait alors un compte à régler avec les militaires - on est à la fin de la guerre d'Algérie et le putsch d'Alger est loin d'être oublié- et les a dépossédés de la recherche au profit de civils²².

Malgré un dossier circonstancié et l'appui du CNRS, la demande de subvention effectuée auprès de Sestier auprès de l'OTAN²³ n'aboutit pas.

La DRME (Délégation Ministérielle pour l'Armement, Direction des recherches et moyens d'essai) prend le relais mais pour un quart seulement du financement, le CNRS continuant à financer le CETA à cinquante pour-cent pour un budget total d'environ un million de francs.

²⁰ Sur l'histoire des réalisations du CETA, voir l'article de Christian Boitet qui prit la direction du GETA, après la disparition de B. Vauquois en 1985 (cf. Boitet 1992).

²¹ On peut douter de la volonté du CNRS, à l'époque, de diffuser largement en France des informations recherches en TA. A la demande de subvention faite par Delavenay en août 60 pour la publication des actes en français du congrès de Teddington, Lejeune répond qu'il n'est pas indispensable qu'une édition soit faite en français puisqu'il y aura déjà une édition en anglais et en russe et que, dans ces conditions, elle ne sera pas rentable. Une partie des actes sera publiée aux PUF sous les auspices de l'ATALA en 1964.

²² Une grande partie des officiers engagés dans des activités de recherche ont alors démissionné de l'armée pour rejoindre l'industrie. Nous remercions René Moreau pour cette information.

²³ Il est probable que cette demande souffrit de la détérioration des relations de la France avec l'OTAN. On sait que De Gaulle manifesta une vive opposition, en décembre 1962, aux accords conclus à Nassau par Kennedy et Macmillan sur le projet de force multilatérale. Cette opposition aboutira en mars 1966 à un double retrait: des forces françaises soustraite à l'autorité de l'OTAN et des forces étrangères qui se trouvaient sur le sol français dans le cadre de l'alliance.

Le CNRS fait alors appel à un certain nombre de ministères afin de financer les vingt-cinq pour-cent restant. Lors d'une réunion regroupant en mai 1962 la direction du CNRS et les représentants des différents ministères concernés, ceux-ci se déclarent favorables aux recherches en TA, mais exposent pour la plupart leur difficulté à disposer des crédits demandés. Quatre-vingt dix mille francs seulement seront réunis à l'issue de cette réunion; le SDEC se montrant le plus généreux en accordant cinquante mille francs.

A l'occasion de ces difficultés de financement, on voit apparaître de façon manifeste, dans les réunions du CNRS chargées de l'évaluation du CETA, des discussions sur la demande sociale en traductions. Ce point qui constituait un élément important du rapport Sestier, préalable à la création du CETA et qui figurait dans ses statuts, à travers la création d'un centre de production de traductions devant assuré l'autofinancement du centre, était peu présent dans les discussions. Lors d'une réunion de juin 1962, notamment, sont évoquées la pertinence de limiter les recherches à la traduction russe - > français et à celle des textes techniques et scientifiques qui n'intéressent pas nécessairement certains ministères, comme par exemple le ministère de la culture. A l'issue de la réunion, ces deux orientations sont toutefois conservées, notamment en raison de l'intérêt que présente la traduction des textes techniques et scientifiques pour les pays en voie de développement.

Limitation des crédits et discussions sur la demande sociale en TA soulèvent crucialement la question des délais dans lesquels il sera possible d'effectuer les premières expériences de TA. Sestier et Vauquois ne sont pas d'accord sur ce point: trois ans suffisent pour Vauquois, alors que pour Sestier, rien n'est possible avant huit ans. Celui-ci laisse entendre qu'il ne pourra pas, avec des délais si courts, continuer à assumer la direction du CETAP. Le directeur du CNRS, Jean Coulomb assure que le CNRS est prêt à jouer le jeu et à continuer à financer le CETA pour moitié et qu'on trouvera les crédits nécessaires de façon à ce que l'affaire se poursuive au moins jusqu'en 1970, date butoir de Sestier.

Un autre facteur enfin est à prendre en compte dans la démission de Sestier: la publication du rapport Bar-Hillel.

Les français semblent n'avoir eu que relativement tardivement connaissance du rapport Bar-Hillel, publié en 1960, et qui évaluait de façon généralement négative les activités et les résultats des centres de TA dans le monde et surtout aux États-Unis²⁴. Lors du Congrès de l'IFIP à Munich en août 1962, où il organisait un symposium intitulé "Modern techniques of language translation", Sestier le mentionnait comme une chose

²⁴ E.Delaveney possédait une version du rapport Bar-Hillel datant de 1959 (cf. archives de l'ATALA).

assez récente, mais la façon dont il en parlait laissait prévoir sa décision²⁵. La direction du CNRS n'ignore pas les positions de Bar-Hillel sur la TA. Dans un courrier du 10 octobre 1962, Jean Coulomb attire l'attention de Michel Lejeune sur le texte de Bar-Hillel dans les actes du Congrès de l'IFIP que vient de lui envoyer Sestier.

Finalement, en octobre 1962, Sestier donne sa démission du CETAP. Dans le rapport d'activité du CNRS 1961-62, les raisons de cette démission sont exposées de la façon suivante: "la construction d'une syntaxe pour une langue naturelle est un problème de recherche fondamentale inséparable de celui de la sémantique; ce problème a été abordé prématurément et exigerait, pour être résolu, de nombreuses recherches préliminaires (de psychologie et de théorie des automates, notamment) auxquelles la structure actuelle du CETAP ne se prête pas."

Après la dissolution du CETAP, Maurice Gross et Yves Gentilhomme rejoignent le Laboratoire de calcul numérique de l'Institut Blaise Pascal. Quant à Jacques Perriault, il rejoint la SAD (Section d'Analyse Documentaire) dirigé par Gardin.

Entretemps, Gross a effectué, grâce à une bourse Unesco sollicitée par Sestier, un stage de recherche aux Etats-Unis d'octobre 1961 à mai 1962 au MIT (sous la direction de Chomsky) et à Harvard. Lors de son séjour, M.Gross a rencontré Bar-Hillel et a discuté avec lui des problèmes que posait la TA et de la médiocrité des résultats obtenus²⁶. A l'IBP, il poursuit des recherches sur études sur la syntaxe formelle des langages de programmation et des langages naturels²⁷, et l'étude de la linguistique structurale et de la syntaxe du français. Il mène des expériences d'analyse et de synthèse sur calculateurs électroniques. Enfin il étudie les correspondances entre langages et traduction²⁸.

Yves Gentilhomme est affecté à l'Institut Blaise Pascal de 1962 à octobre 1965, date à laquelle il est nommé assistant à la Faculté des lettres de Besançon. En thèse avec Martinet, il étudie la morphologie russe en vue d'une formalisation. Par ailleurs il enseigne le russe à l'EPTR (Enseignement Préparatoire aux Techniques de la Recherche) du CNRS et publie en 1963 un manuel de russe à l'usage des scientifiques. En mai 1965, il obtient une mission d'un mois aux USA, où il visite les centres de TA²⁹.

²⁵ Nous remercions André Lentin pour cette information.

²⁶ Je remercie Maurice Gross de m'avoir fourni cette information.

²⁷ Ces recherches aboutiront notamment à la rédaction avec André Lentin des *Notions sur les grammaires formelles* publié en 1967.

²⁸ Comme Maurice Gross, David Cohen fait partie du groupe 'linguistique formalisée' de l'IBP, où il élabore une analyse automatique de la morphologie et de la syntaxe de la langue arabe.

²⁹ C'est également en 1965 qu'il devient secrétaire administratif de l'ATALA, avant d'en être le vice-président avec Georges Gougenheim de 1971 à 1975 et président de 1978 à 1980.

Ne reste donc plus que l'ATALA et le groupe de Grenoble, qui, tout en restant rattaché à l'Institut Blaise Pascal, devient laboratoire propre du CNRS en 1963, en gardant le nom de CETA. Entre-temps est créé le groupe de Nancy.

6. Le Groupe de TA de Nancy

Le Groupe de TA de Nancy est créé à la suite d'une réunion le 25 mai 1960 à la Faculté des Lettres de Nancy, organisée par le Doyen Schneider et A. Culioli, alors Professeur à la Faculté des Lettres de Nancy, à laquelle participent, outre Bernard Pottier et Guy Bourquin, le président de l'ATALA E. Delavenay, A. Martinet, et M. Legras, directeur de l'Institut de Calcul Automatique de l'Université qui met à la disposition du groupe une machine IBM 650.

Au départ, le groupe comprend deux linguistes, Bernard Pottier, professeur à la Faculté des lettres de Strasbourg et Guy Bourquin, chargé d'enseignement à la Faculté des Lettres de Nancy.

Avec le soutien du Doyen Schneider et du Recteur Imbs, le groupe de Nancy obtient d'être examiné par le CNRS lors du comité de direction du CETA dès janvier 1962, et en juillet 63 il obtient deux collaborateurs techniques supplémentaires. En 1969, il devient le CRAL (Centre de Recherche et Applications Linguistiques de Nancy), associé au CNRS, dont une section s'occupe du traitement automatique des textes sémitiques, paléographie automatique et analyse documentaire.

Le groupe de Nancy se donne comme langue source l'anglais et comme langues cibles le français et l'espagnol, choix, dit-il, qui se justifie par un intérêt pratique, la traduction d'ouvrages scientifiques. Sa méthode est fondée sur des critères linguistiques, et s'attache au début à la délimitation et à l'analyse morpho-syntaxique des syntagmes ainsi qu'à la définition de filtres syntaxiques destinés à éliminer les ambiguïtés. Elle se présente comme un moyen terme entre la méthode empirique (la traduction mot à mot) et la formalisation mathématique. Une de ses préoccupations essentielles concerne l'étude des problèmes linguistiques posés par la TA sur lesquels plusieurs thèses sont engagées très tôt.

7. De la crise de l'ATALA de 1965 au rapport Gross de 1967

A partir de fin 1962, Vauquois monte en puissance dans les instances représentatives. En 1963, il est membre de la section 22 'Linguistique générale, langues moderne et littérature comparée'. Puis il sera membre de la section 28 'Linguistique générale, langues et littératures étrangères' en 1969. Dès 1965, il est vice-président de l'ATALA. Il en sera le président de 1966 à 1971. Il sera un des fondateurs, également en 1965, de l'ICCL (International Committee on Computational Linguistics).

Le CETA continue de prospérer. Avec un budget d'un million de francs et 20 collaborateurs en 1963, il atteint les 35 personnes début 1965. Ce budget de 1 million

sera augmenté en 1966. En 1964, un budget spécial est voté pour la construction du bâtiment du CETA sur le campus de St Martin d'Hères, bâtiment qui sera inauguré en août 1967.

7.1. La crise de 1965

La crise de 1965 qui affectera profondément l'ATALA, ne semble pas ébranler le CETA.

Le déclin de la Traduction Automatique, amorcé depuis 1960 aux États-Unis suite à la publication du rapport Bar-Hillel, est accentué avec la constitution de l'ALPAC (Automatic Language Processing Advisory Committee) en 1964 et la publication de son rapport en 1966. Ce déclin déclenche une crise au sein de l'ATALA en 1965.

Celle-ci ne doit sa survie qu'à l'ingénieur général Nicolau³⁰, membre fondateur et vice-président de l'ATALA, lors de l'Assemblée Générale du 15 mai 1965. Selon lui, les échecs de la traduction automatique ne doivent pas entraîner la disparition de l'ATALA ni ralentir ses activités; celle-ci doit s'orienter résolument vers l'une de ses missions, le développement de la linguistique appliquée et donner moins d'importance à la TA. Il remercie les "croyants" de leur aide et devient président de l'ATALA. Vauquois en devient le vice-président.

A l'issue de cette crise, l'ATALA et sa revue changeront de nom: elle deviendra *l'Association pour le Traitement Automatique des Langues* et son bulletin *La Traduction Automatique*, deviendra *TA Informations, Revue internationale des applications de l'automatique au langage*.

Les thèmes se diversifient et laissent moins de place à la TA en tant que telle; celle-ci apparaît sous la rubrique plus vaste de *automatique documentaire, linguistique mathématique et traduction automatique*.

Elle ouvre ses colonnes aux nouvelles de l'AILA qui occupera un bon tiers de chaque livraison. Désormais, le domaine de la linguistique appliquée ne couvre plus seulement les questions d'enseignement des langues, mais celles de l'informatique documentaire et du traitement automatique des langues

Malgré cette crise, selon le rapport de conjoncture du CNRS 1963-64, les sections de linguistique du CNRS maintiennent leur soutien prudent à la TA. Ainsi la section 23 'linguistique française et études littéraires' mentionne dans ses recommandations l'intérêt de la linguistique appliquée dont fait partie la TA qui a notamment pour aboutissement une analyse plus poussée de la structure actuelle du français encore très imparfaitement connue.

³⁰ En janvier 65, René de Possel, directeur de l'Institut Blaise Pascal et président du Conseil scientifique du CETA, propose que le Général Nicolau, ainsi d'ailleurs que M.P. Schützenberger deviennent membres du conseil scientifique du CETA.

La section de mathématiques appliquées, quant à elle, accuse la crise mais de façon très modérée. Elle parle d'une phase de reflux, mais précise que les études faites en vue de la TA conduiront un jour à des résultats intéressants à condition d'être menées en profondeur et que les études en cours doivent être poursuivies avec toutes les ressources actuelles.

Dans ses recommandations, la section ne semble pas affectée par la démission de Sestier et encourage la collaboration entre linguistes et logiciens pour mener à bien une "étude minutieuse et patiente des langues naturelles". Par ailleurs, l'idée de construire une machine dédiée à la traduction est loin d'être abandonnée et la section préconise des études de logique de circuits et d'organes de machine adaptées à des fins de traduction, de même que des études de programmation dans le domaine non numérique visant à la création et compilation de langages adaptés

Autrement dit les problèmes de TA relèvent à la fois de la linguistique, de la logique, de l'électronique et de la programmation.

Concernant plus spécifiquement les activités du CETA, le conseil scientifique de décembre 1966 recommande de diversifier les domaines d'applications des méthodes élaborées par le centre. La TA n'est plus à considérer comme le but premier et doit apparaître comme un produit dérivé d'études linguistiques et logiques plus générales.

On notera qu'à la faveur de la crise, la question de la demande sociale surgit de nouveau et le conseil scientifique évoque une enquête en cours, celle du BIPE (Bureau d'Information et de Prévisions Economiques) destinée à évaluer les besoins en TA sur le plan national.

Malgré la crise le CETA tient donc bon. Ses crédits sont augmentés et ses locaux sont agrandis. De Possel, demande au nom du Conseil Scientifique réuni le 15 décembre 1966, les félicitations pour le travail accompli par Vauquois et évoque les vastes possibilités, autres que la TA, offertes par les efforts accomplis par le CETA. Enfin la seconde conférence Internationale sur le TAL (Traitement Automatique des Langues) se tiendra en août 1967 au CETA qui y présentera sa première expérience de traduction sur ordinateur (le premier test aura lieu en juin 1967).

En ce qui concerne les recherches, c'est la mise au point d'un langage d'écriture des règles syntaxiques à partir de fin 1963, puis à partir de 1965 la mise au point du langage pivot, qui constituent les recherches les plus originales du CETA qui opta ainsi pour une méthode de TA utilisant une langue intermédiaire. Le langage pivot est un modèle syntaxico-sémantique qui assure l'indépendance des phases d'analyse et de synthèse dans le processus de traduction.

En 1970, comme pour la plupart des autres groupes, la TA est considérée par le CETA comme le transfert de sens d'un texte écrit d'une langue source à une langue cible. Le niveau du sens ne pouvant encore ni être défini formellement ni atteint par traitement

automatique, il faut trouver un niveau intermédiaire qui donne une approximation suffisante. C'est ce niveau intermédiaire que tente de formaliser le langage pivot.

7.2. Le rapport Gross, septembre 1967

Cette confiance dans le CETA est à peine ébranlée par le rapport Gross de septembre 1967 adressé à la direction du CNRS. Ce rapport, outre qu'il promeut le français comme sujet d'étude principal et la phonologie et la syntaxe comme seules scientifiques, est une critique sévère du TLF (Trésor de la Langue Française) et du CETA qui, selon ses propres termes, déconsidèrent la France aux yeux de la communauté scientifique internationale. Il se termine par une présentation des travaux du LADL (Laboratoire d'Automatique Documentaire et Linguistique) créé par Gross et qui faisait encore partie de l'Institut Blaise Pascal.

Dans ce rapport, M.Gross reproche tout d'abord au CETA son erreur d'orientation. Reprenant les positions énoncées par Sestier en 1959, il définit la TA comme ayant pour objectif "de produire sur des bases commerciales un produit fini utilisable" (p.2). En cela, la TA s'oppose à la linguistique mathématique (statistiques appliquées au langage), à l'analyse des structures des langages artificiels et des langues naturelles, et aux aides mécaniques à la traduction. La TA est une opération industrielle en tout ou rien, une traduction rudimentaire étant totalement inutile. Or malgré la médiocrité des résultats, elle a suscité l'enthousiasme des commanditaires. Par ailleurs le CETA commet une erreur commune aux pays occidentaux, USA et en Europe de l'Ouest, qui consiste à confondre TA et "problème de développement, c'est-à-dire d'adaptation de la linguistique à l'informatique et vice et versa". On retrouve la position de Sestier lorsqu'il poursuit: "La linguistique traditionnelle orientée vers la philologie, l'enseignement des langues et les prescriptions normatives n'a jamais cherché à atteindre un niveau de cohérence, et de précision dans la confection des modèles, tel que l'on puisse envisager son application. Les théories linguistiques modernes (théories transformationnelles de Z.S. Harris et de N.Chomsky) sont trop récentes pour qu'on puisse envisager leur usage pour des réalisations de type industriel, même à un niveau expérimental. D'ailleurs la simple idée d'utiliser le langage comme outil d'analyse systématique du langage lui-même est extrêmement moderne et fort peu intuitive. Elle est du même type que celle qui a conduit à la logique mathématique." (p.3)

Gross s'étonne de l'état de santé florissant du CETA alors que, suite au rapport de l'ALPAC, tous les centres américains ont fermé ou se sont reconvertis et qu'il ne reste au monde que trois ou quatre équipes de moindre importance qui font de la TA. En 1967, dit-il, une activité de TA ne peut se justifier que si elle présente des caractères qui la distingue de celle de tous les centres qui ont disparu. Ce n'est pas le cas du CETA qui, selon lui, n'a rien publié d'original que ce soit en linguistique formelle, en logique mathématique ou en algorithmique.

A l'appui de ces critiques, Gross cite l'exemple de traduction présenté en août 67 à Grenoble lors de la seconde conférence internationale sur le traitement automatique des langues, qui, selon lui, présente tous les défauts habituels: contresens, ambiguïtés, phrases non comprises, phrases indétectablement erronées. Par ailleurs l'ambiguïté relevée par le CETA, à savoir la détermination de l'adjectif *enrhumée* dans le syntagme *la directrice de société enrhumée* n'existe pas pour un locuteur humain. Elle n'existe pour la machine qu'à cause de l'extrême pauvreté de la grammaire du CETA qui ne comporte pas la distinction humain/ non-humain. C'est un cas typique de pseudo-ambiguïté au niveau algorithmique, dit Gross, qui rend la traduction d'une phrase impossible ou inutilisable.

Les raisons de l'absence de résultats théoriques du CETA, sont claires selon Gross: il n'existe pas d'enseignement spécifique de la linguistique en France. Alors qu'il existe depuis un certain nombre d'années aux USA, aucun membre du CETA n'a effectué de formation sur place. Gross cite à ce propos Chomsky, Harris, Hockett et Joos. Par ailleurs, la langue source étant le russe, Gross s'étonne qu'aucun linguiste russisant ne soit agrégé ou ne soit allé étudier à Oxford où se trouve le principal centre d'étude de la grammaire russe dirigé par le professeur Unbegaun.

Aucun chercheur CNRS n'est attaché à l'équipe, ce qui confirme, selon Gross, l'orientation vers le seul développement des travaux en TA, orientation qui fut celle des travaux américains maintenant abandonnés. De plus il souligne l'absence de chercheurs en linguistique dans l'élaboration de la 'langue pivot'.

Gross conclut son rapport en s'interrogeant sur la rentabilité du CETA et l'opportunité de la poursuite de ses activités. Il est intéressant de voir que le rapport Gross remet au jour la contradiction soulevée par Sestier entre demande sociale de traductions en série face à des outils théoriques impuissants à réaliser une telle tâche.

Le rapport Gross ne semble pas avoir eu beaucoup d'effet. Lors du conseil scientifique³¹ qui suivit, en janvier 1968, les membres du conseil examinent et discutent les traductions des deux premiers articles scientifiques du corpus, soit deux cents phrases. Vauquois précise qu'un des problèmes essentiels concerne la stabilisation de la grammaire lorsque le corpus est augmenté de nouveaux textes. M. Zarrouati de la DRME, qui était aussi un des destinataires du rapport Gross, demande à Vauquois de faire le point sur les progrès réalisés entre le mois d'août 1967, date du congrès de TAL où a été présenté le premier essai de TA, et décembre 1967. Dans sa réponse, Vauquois fait le point sur les différents développements, notamment l'accent mis sur la grammaire de synthèse du français, et conclut par l'annonce d'une prochaine expérience

³¹ Étaient présents à ce conseil présidé par R. de Possel: Culioli, Delavenay, Fourquet, Gardin, Genuys, Gsell, Jobert, Kuntzmann, Vauquois et Zarrouati. Étaient absents ou excusés: Gougenheim, Lejeune, Lentin, Martinet, Nicolau, Pottier, Quemada et Schützenberger.

prévue au printemps suivant sur un corpus différent. Réponse qui semble satisfaire Zarrouati qui suggère, si cette nouvelle expérience est concluante, de créer un centre de traductions opérationnel. A une question de Gardin demandant s'il n'est pas trop tôt pour passer de la recherche au développement, Zarrouati répond en invoquant la forte demande en traductions qui pousse à envisager une phase de production. Quant aux questions théoriques sur la validité du modèle du langage pivot, évoquées aussi par Gardin, elles ne semblent guère devoir être approfondies. Le Conseil se prononce pour une poursuite des travaux tout en précisant que soit menés en priorité la mise au point et l'accroissement du système de traduction russe-français.

Lors de cette réunion, un second courant au sein du CETA est explicitement encouragé par le Conseil scientifique. Il s'agit de l'équipe de recherche sémantique dirigée par Jacques Rouault, engagée dans l'étude de la formalisation des modalités, de l'aspect et des temps verbaux.

Malgré cette prospérité apparente du CETA et le soutien du CNRS, on peut toutefois penser que le rapport de l'ALPAC, relayé par le rapport Gross, n'a pas été sans effets.

En mai 1968 l'équilibre interne du CETA sera profondément affecté; ce qui aboutira notamment à la scission des personnels en deux groupes et à la création en janvier 1971 de deux équipes associées au CNRS: le GETA d'une part, au sein de l'institut de recherche en mathématiques avancées, toujours dirigé par Vauquois, poursuit les activités de l'ancien CETA. L'équipe de traitement automatique des langues est créée par le CNRS à l'université des sciences sociales de Grenoble sous la direction de Jacques Rouault³².

On peut avancer plusieurs raisons pour expliquer la scission du CETA.

Les qualités scientifiques de Vauquois, reconnues par tous, et qui lui valaient l'estime de ses collaborateurs, ne suffisaient pas à assurer une gestion du personnel suffisamment bonne pour maintenir un climat de travail exempt d'amertumes et de rivalités au sein de l'équipe. Mais les raisons sont essentiellement scientifiques. Il n'y avait pas de chercheurs dans l'équipe, et surtout pas de chercheurs en linguistique. Rouault était maître de conférence en mathématiques appliquées. Ainsi bien que l'étude des langages formels tienne une grande place dans les recherches du CETA, Vauquois privilégiait l'informatique à la linguistique.

Enfin le courant culiolien qui se développait à l'intérieur du CETA sous la direction de Rouault, en connexion avec l'École d'Aussois animée notamment par Grize, Pêcheux, Culioli, Bresson, était peu impliqué dans les recherches de TA. Dès 1966, des

³² Cette équipe ne sera pas reconduite en 1976.

recherches sur la formalisation sémantique et la notion de ‘lexis’ de Culioli³³ avaient engagées au sein du CETA autour de Jacques Rouault.

L’équipe, une fois créée, se consacrera à la reconnaissance automatique du français fondée sur une analyse linguistique approfondie³⁴. Deux autres projets, portant l’un sur la logique du langage et l’autre sur la théorie du discours sont alors au stade exploratoire.

Le domaine de recherche en TAL, nom de la nouvelle équipe, apparaît pour la première fois³⁵ dans la partie du rapport de conjoncture du CNRS de 1969, consacrée à la section de ‘Linguistique générale, langues et littératures étrangères’, dont je le rappelle fait partie Vauquois ainsi que Culioli, Delavenay, et Pottier. Sont concernés le CETA, le LADL, le groupe de TA de Nancy devenu le CRAL (Centre de Recherche et d’Applications Linguistiques) de Nancy et le CADA (Centre d’Analyse Documentaire pour l’Archéologie) dirigé par Gardin. Sont concernées enfin les associations comme l’ATALA et l’AFLA.

Le terme de TAL disparaîtra des rapports du CNRS en 1972 au profit de linguistique informatique.

Conclusion

En conclusion, nous voudrions revenir sur la question posée au commencement de cet article, à savoir comment expliquer la prospérité du CETA en 1967 malgré le rapport de l’ALPAC et suivant un mouvement inverse à l’évolution mondiale.

En fait il faut pondérer cette impression de prospérité: le CETA se porte très bien en 1967, mais moins bien en 1971, où il sera privé d’une partie de ses forces, chercheurs et moyens, et perd son statut privilégié de laboratoire propre du CNRS. Donc le rapport Gross n’a sans doute pas été sans effet et l’affirmation du courant culiolien s’ancrant dans le TAL n’a pas été négligeable.

³³ C’est ainsi que trois jeunes linguistes, Catherine Fuchs, alors en thèse avec Culioli, Sophie Fischer et Hélène Pauchard, font leur mémoire de fin de stage de l’EPRASS, intitulé ‘Modèle réduit d’un analyseur syntaxique du français’ dans le cadre du CETA à Pâques 1968. L’EPRASS (Enseignement Préparatoire à la Recherche Approfondie en Sciences Sociales) était un DEA expérimental en deux ans, répondant à une demande de l’université concernant la première année de thèse: celle-ci devait être une formation à la recherche sanctionnée par un diplôme. Il a été expérimenté dans les disciplines, telle la linguistique, où il n’y avait pas d’agrégation.

³⁴ Elle comprendra des ingénieurs comme Claude Del Vigna, des informaticiens comme Mireille Dupraz ou Vincent Meissonnier, et continuera à bénéficier de la collaboration de Catherine Fuchs, entrée comme chargée de recherche au CNRS en 1969 dans l’équipe de Culioli.

³⁵ C’est l’ATALA qui semble avoir utilisé pour la première fois le terme à l’issue de la crise de 1965. Antérieurement *la* TAL désignait la Traduction Automatique des Langues (cf. rapport d’activité CNRS de Sestier du 22 mars 1962).

Mais cela s'est fait progressivement et non brutalement comme aux États-Unis. On peut invoquer le fait que l'instance d'évaluation et le bailleur de fonds qu'est le CNRS assure une certaine stabilité aux équipes. Il faut aussi prendre en compte un certain nombre d'aspects propres au CNRS: présence des intéressés dans les instances d'évaluation, façon dont est prise en compte la demande sociale, l'évaluation des modèles théoriques et l'évaluation des résultats. Dans le cas de la TA, qui doit viser avant tout une production industrielle, les choses sont éminemment ambiguës.

Par ailleurs on peut supposer cette stabilité a permis au CETA, devenu GETA, de prendre un second souffle et de réaliser dès le début des années 70, Ariane-78, un système multi-niveaux considéré comme un des plus performants de l'époque.

- **Sources:**

- **Fonds documentaire du CNRS de Gif-sur-Yvette**

- rapports d'activité annuels du CNRS (1950-1975)
- rapports de conjoncture du CNRS (1959-1974)

- **Fonds d'archives du CNRS de Gif-sur-Yvette**

- dossiers scientifiques CNRS 910024 DPC
- arriérés d'archives (1955-1963)

- **Fonds d'archives du Centre des Archives Contemporaines de Fontainebleau**

- archives historiques F 780 309 (53) et (54)

- **Archives de l'ATALA**

- **Rapports**

- Gross M. 1967, *Notes sur certains aspects des recherches en linguistique au CNRS*, rapport au directeur général du CNRS du 26 septembre 1967.
- Holmstrom J.E., 1951, *Rapport sur les dictionnaires scientifiques et techniques multilingues*, Unesco Paris
- *Language and Machines. Computers in translation and linguistics*. 1966. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council.
- Sestier A., 1959, *Comment doit être organisé à l'échelle française l'effort pour la TA* (rapport du 23-11-59).
- Sestier A., 1961a, *Justification d'une demande de subvention à l'OTAN*, version 1: juin 1961; version 2: 8 octobre 1961.
- Sestier A., 1961b, *L'automatisation de l'analyse syntaxique*, Document du CETAP, note n°13, 19 octobre 1961
- UNESCO, 1953, *Report on scientific and technical translating and related problems* [WS/073.119]
- Vauquois B., 1961, *Étude sur le choix du matériel destiné à la traduction automatique*, 13 janvier 1961

- **Bibliographie**

- Actes du colloque *De Bourbaki à la Machine à Lire, Journée d'hommage à René de Possel*, ENS-Ulm, 16 novembre 1994

- Archaimbault S. et Léon J., 1997 "La langue intermédiaire dans la Traduction Automatique en URSS (1954-1960). Filiations et modèles", *Histoire Épistémologie Langage* 19-2:105-132.
- Boitet Ch., 1992, "TA et TAO à Grenoble ... 32 ans déjà!" *TAL*, vol 33 n°1-2,:45-84.
- Boucher Henri, 1988, "L'informatique dans la défense", *Actes du colloque sur l'Histoire de l'Informatique en France*, Grenoble mai 1988, :83-102.
- Cahiers pour l'histoire du CNRS (1939-89)
- Chevalier J-C., Encrevé P., (1984), "La création de revues dans les années 60. Matériaux pour l'histoire récente de la linguistique en France", *Langue Française* n°63.
- Chiss J-L. et Puech Christian (eds), 1998, 'La linguistique comme discipline en France', *Langue Française* n°117, février 1998.
- Delavenay E. et Delavenay K., 1960a, *Bibliographie de la traduction automatique*, Mouton § co. 's-Gravenhage.
- Delavenay E., 1960b, *La machine à traduire* Que-sais-je?
- Delavenay E., 1992, *Témoignage. D'un village savoyard au village mondial*, diffusion Edisud, Aix-en-Provence.
- Lentin A., 1992, "Naissance et premiers pas de l'ATALA: quelques souvenirs et quelques réflexions" *Traitement Automatique des Langues*, N° spécial trentenaire, 1992, vol.33, n° 1-2 :7-24.
- Léon J., 1992, "De la traduction automatique à la linguistique computationnelle. Contribution à une chronologie des années 1959-1965", *Traitement Automatique des Langues* , N° spécial trentenaire, 1992, vol.33, n° 1-2 :25-44.
- Léon J., 1998, "Langues auxiliaires, traductions, et modèles de traduction automatique (1950-1970)", XI. *International Colloquium of the Studienkreis der Geschichte der Sprachwissenschaft*, Leuven, 2-4 juillet 1998.
- Mounier-Kuhn P., 1994, "L'Institut Blaise Pascal du CNRS de 1946 à 1969", actes de *De Bourbaki à la Machine à Lire, Journée d'hommage à René de Possel*, ENS-Ulm, 16 novembre 1994: 13-30.
- Nolin L., 1998, "Des caves de l'Institut Henri Poincaré à la terre promise de la rue du Maroc" rencontre scientifique en l'honneur d'André Lentin 23 février 1996, *Mathématiques, Informatique et Sciences Humaines*, n°141, :5-10
- Ramunni J., 1989, *La physique du calcul, Histoire de l'ordinateur*, Hachette, Paris, 287pp.
- Ramunni Girolamo, 1991, "La non-construction du premier calculateur électronique au CNRS (I)", *Micro-bulletin* n°42, decembre 1991
- Ramunni Girolamo, 1992, "La non-construction du premier calculateur électronique au CNRS (I) et (II)", *Micro-bulletin* n°-43-44-45, février-mars 1992
- Sestier A., 1959, "La traduction automatique des textes écrits scientifiques et techniques d'un langage dans un autre", *Ingénieurs et Techniciens*, partie I et II.