

Automatisation-mathématisation de la linguistique en France dans les années 1960. Un cas de réception externe

Léon, Jacqueline

UMR7597, Histoire des Théories Linguistiques, CNRS, Université Paris-Diderot, Paris 7
jleon@linguist.jussieu.fr

L'automatisation-mathématisation de la linguistique constitue le second tournant de la mathématisation de la linguistique au 20^e siècle. La première mathématisation est le tournant logico-mathématique des années 1930 ; le second tournant est le domaine aux confins de l'analyse syntaxique, des langages formels et de la programmation qui prendra le nom, aux Etats-Unis, de « computational linguistics » et qui se constituera en même temps et à partir des mêmes fondements que les formalismes grammaticaux¹.

Cette base commune, c'est la traduction automatique, conçue comme technologie de guerre aux Etats-Unis dans les années 1950. La linguistique computationnelle, cette « nouvelle linguistique », créée en dehors de la linguistique, va devoir s'intégrer dans les sciences du langage dans les années 1960. Les modes d'intégration sont divers selon les traditions intellectuelles et les institutions en présence dans l'après-guerre.

Pour la France, nous proposons de parler de réception externe de 'l'automatisation-mathématisation' plutôt que d'intégration. La réception externe, par opposition à l'intégration, est un processus qui ne s'appuie sur aucun ancrage dans la tradition ni dans l'horizon de rétrospection du champ de recherche qui va l'accueillir. Plusieurs facteurs contribuent à expliquer cette particularité française : l'absence de configuration propre de sciences pour la guerre (Dahan et Pestre, 2004), l'absence d'ancrage dans la logique mathématique et la philosophie du langage, la concurrence entre linguistique computationnelle venant des Etats-Unis et linguistique mathématique venant d'URSS, le retard français en matière d'équipement informatique et d'expérimentations de traduction automatique (désormais TA), la quasi-absence de lieux de formation spécifiques dans les universités. Ces divers facteurs ont eu pour conséquence que l'intégration de la nouvelle mathématisation de la linguistique se soit faite de façon totalement externe. Nous examinerons ce mode de réception à travers un certain nombre de questions :

- comment les Français, si étrangers à l'idée même 'd'automatisation-mathématisation', vont enclencher le processus ?
- est-ce que l'automatisation-mathématisation va être uniquement liée à la TA et s'effectuer au sein des centres créés à cet effet, ou bien mobiliser d'autres disciplines ? Des opérateurs de passage (Chiss et Puech, 1999), hommes ou institutions ne seront-ils pas nécessaires ?
- comment les linguistes vont-ils réagir à cette imposition d'un nouvel « horizon de rétrospection » (Auroux 1987, 2006) venu d'ailleurs et qui leur est totalement extérieur ?
- quels effets sur la profonde restructuration des sciences du langage des années 1960 ?

1 Tradition française et horizon de rétrospection des sciences de la guerre

La France reste complètement à l'écart du mouvement logico-formaliste des années 1930, où se trouve ancrée la première mathématisation du langage du 20^e siècle. Ce mouvement, à l'intersection du formalisme, du positivisme logique du Cercle de Vienne et de la philosophie analytique du langage de Cambridge, comprend les figures centrales de Gottlob Frege, Rudolph Carnap, Bertrand Russell et Ludwig Wittgenstein. Dans un tel cadre, où la logique formelle unifie, structure et codifie les sciences, les

mathématiques deviennent un langage formel comme les autres et marquent la première mathématisation du langage.

La situation en France peut s'expliquer par la désaffection de l'enseignement de la logique due notamment au destin tragique des logiciens français (Jean Cavaillès ou Jacques Herbrand), par l'attraction qu'exerce sur les philosophes français la tradition philosophique allemande nietzschéo-heideggerienne, enfin par la radicalité du mouvement structuraliste (voir sur ces questions Dosse, 1991).

A la fin des années 1950, au sein de la Société de Linguistique de Paris (désormais SLP) toute puissante, les travaux structuralistes américains ne suscitent pas beaucoup d'intérêt (Chevalier, 1990). Quand ils sont recensés dans le *BSL*, ce qui arrive rarement, ils sont souvent sévèrement critiqués, comme par exemple *Methods in Structural Linguistics* de Z.S. Harris recensé par Jean Cantineau en 1954. Le compte rendu de R-L Wagner dans le *Journal de psychologie normale et pathologique* (1954) est à peine plus positif. Il souligne le caractère sec, limité et déconcertant des analyses distributionnelles qui rendent les travaux de Harris, et des structuralistes américains en général, difficilement accessibles aux linguistes français. La situation s'améliore un peu une fois Georges Mounin devenu membre de la SLP en 1958. C'est lui qui écrira le compte rendu de *Syntactic Structures ...* en 1961. Dans le volume de 1966 du *BSL*, André Mirambel, alors membre du bureau de la SLP, consacre une petite page au compte rendu du neuvième Congrès International des Linguistes de 1962. Bien que ce congrès soit considéré (Murray, 1994) comme le lieu où Chomsky se serait fait véritablement connaître au niveau international, Mirambel ne mentionne ni son nom ni ses travaux. Il se contente de noter l'importance accordée à la linguistique mathématique, aux problèmes de mesure, à la phonologie, à la stylistique et aux problèmes de structure (fait-il allusion ici aux travaux de Chomsky ?). Bien qu'insistant à plusieurs reprises sur l'importance du renouvellement des méthodes, sans d'ailleurs expliciter en quoi celui-ci consiste, il n'en saisit manifestement pas les enjeux.

Les travaux en traduction automatique sont très peu représentés, à l'exception toutefois de la recension en 1957 par J. Porte du premier ouvrage collectif de 1955 publié aux Etats-Unis (Locke & Booth, 1955). Dans le *Journal de Psychologie normale et pathologique* en 1959, le compte rendu de cet ouvrage par P. Guiraud, insistant sur les aspects techniques de la TA, n'incite guère les linguistes à s'intéresser à la question. Il mentionne la parution de « 14 essais faits par un groupe de techniciens qui, depuis une dizaine d'années déjà, travaillent à la réalisation d'une machine à traduire. Les problèmes sont à la fois d'ordre mécanique et d'ordre linguistique (élaboration du dictionnaire) » (Guiraud, 1959 : 482).

En revanche, les travaux issus des pays du bloc soviétique sont systématiquement recensés. Dès 1955, René Lhermitte rédige des comptes rendus réguliers des revues russes *Voprosy Jazykoznanija* et *Izvestija Nauk SSSR* où figurent les travaux soviétiques en linguistique mathématique. Quant à Marc Vey, il recense les revues tchèques dont *The Prague Bulletin of Mathematical Linguistics*. Ainsi, au début des années 1960, les linguistes disposent d'une information bien meilleure sur les travaux soviétiques en linguistique mathématique que sur les travaux américains, que ce soient les analyses distributionnelles ou les grammaires formelles. Ce qui explique, en partie, comme on le verra plus loin, qu'au sein de la linguistique française il y aura d'autres formes de mathématisation que la conjonction syntaxe-formalismes-programmation.

Enfin, la configuration des sciences de la guerre, creuset de la seconde mathématisation du langage, n'est que très partiellement représentée en France. Avec la quasi-inexistence de la logique mathématique, les grandes théories unificatrices, telles que la cybernétique, la théorie de la communication et la théorie de l'information, conçues et élaborées dans cette configuration ne trouvent pas l'assise nécessaire pour leur importation.

L'informatique, au coeur du tournant de l'automatisation-mathématisation, est très en retard en France. Alors qu'aux Etats-Unis, l'ordinateur est un des dispositifs technologiques qui illustre le mieux la réussite de cette nouvelle alliance entre sciences et ingénierie des sciences de la guerre. En 1944 à Harvard, est construit le premier calculateur numérique, Mark I, dont la supériorité sur les calculateurs analogiques vient d'être montrée. Or, en France dans les années 1950, il y a encore une nette prédominance du calcul analogique sur le calcul numérique. 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. Après un échec cuisant de la construction d'une machine française, le CNRS décide d'équiper en 1955 l'Institut Blaise Pascal d'une machine anglaise, une Elliott 402. La première machine française est donc anglaise (Ramunni, 1989).

De façon générale, l'armée et l'industrie françaises ne manifestent qu'un intérêt frileux pour les calculatrices électroniques. De son côté, l'université fait preuve d'une très grande hostilité à l'égard des machines et du traitement de l'information. Outre le retard pour la construction du matériel, ce désintérêt a pour conséquence une pénurie générale de personnel spécialisé.

2 Contexte d'apparition de la seconde mathématisation en France

Aux Etats-Unis et en Grande-Bretagne, la seconde mathématisation se constitue dans les équipes qui font de la traduction automatique. La TA, conçue par Warren Weaver, personnage clé des sciences de la guerre, est une véritable technologie de guerre (froide), chargée de produire des traductions en série d'articles scientifiques soviétiques pour des raisons stratégiques et économiques. En France, le tournant de cette seconde mathématisation va s'effectuer aussi avec les recherches en TA mais du fait du retard d'une dizaine d'années et de l'absence d'ancrage spécifique, ce ne sera pas le seul lieu ni le seul mode de réception de l'automatisation-mathématisation. Il n'y aura pas comme aux Etats-Unis le cheminement de l'analyse syntaxique automatique qui va s'imposer comme une nouvelle discipline à la conjonction de l'horizon de rétrospection des sciences de la guerre, notamment de la logique mathématique et de la théorie des automates, et de la linguistique. La posture des linguistes par rapport à ce nouveau champ va également être très différente.

Le fait que la TA en France apparaisse en 1959 n'est pas un hasard². Les centres de TA sont créés dans le sillage de l'avènement de l'ère gaullienne du développement d'une véritable politique de la recherche scientifique. Le CNRS connaît un développement sans précédent et deviendra une puissante « machine » de recherche³. D'autres organismes sont créés, comparables au modèle américain, comme la DRME (Direction des Recherches et Moyens d'Essais) où les militaires promeuvent une recherche intégrée entre ingénieurs et universitaires.

Mais - sans doute en raison du faible intérêt pour l'informatique évoqué plus haut - ce n'est ni un universitaire ni un scientifique qui impulse la TA en France. C'est Emile Delavenay (1905-2003) qui attire dès 1957 l'attention sur les travaux en TA. Il 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 assiste le linguiste Michel Lejeune, alors directeur adjoint des lettres et sciences humaines du CNRS, participent les mathématiciens Georges Guilbaud, Benoît Mandelbrot, et Marcel-Paul Schützenberger - ces deux derniers joueront un rôle crucial, on le verra, dans le processus de mathématisation. Emile Delavenay, du fait de ses fonctions de directeur du service des documents et des publications de l'UNESCO, et d'ancien responsable des services de traduction et d'édition à l'ONU, connaît bien les travaux de TA sur le plan international et crée un groupe de réflexion et d'information qui deviendra en avril 1959 une association, l'ATALA, à laquelle sera associée une revue, *La Traduction automatique*, créée en 1960. La vocation de l'ATALA est d'être plutôt un forum de discussion sur la traduction automatique, la documentation automatique et la linguistique appliquée qu'un organisme de recherches à proprement parler. Ainsi, le premier centre de TA est une société savante, et la pénétration du domaine en France a nécessité l'intervention d'un 'passeur-entremetteur' externe. Delavenay n'est ni un ingénieur, ni véritablement un universitaire (il est quand même normalien) mais c'est un homme de réseau et de frontières.

Le CNRS, grâce notamment à l'appui de Lejeune, crée en décembre 1959, au sein du Laboratoire de calcul numérique de l'Institut Blaise Pascal, le CETA (Centre d'Etude en Traduction Automatique), avec deux sections, l'une à Paris dirigé par Aimé Sestier, l'autre à Grenoble dirigé par Bernard Vauquois.

Alors qu'aux Etats-Unis, nombre de centres de TA sont créés dans les universités, au sein des départements de langues, en France ils sont créés au CNRS dans un centre de mathématique appliquée (à l'exception du centre de Nancy, créé en 1960 à l'université de Nancy avec les linguistes Bernard Pottier

et Guy Bourquin). Les spécialistes de langues dans les centres de TA sont des diplômés de langues, en particulier de russe, mais ce ne sont pas des universitaires. La TA se présente avant tout comme un problème pour les calculatrices électroniques, et ce sont donc les ingénieurs et les spécialistes de mathématiques appliquées qui prennent en main le domaine.

Toutefois, contrairement à ce que laissait présager leur faible intérêt pour la linguistique américaine, les linguistes ne se désintéressent pas de la TA. Ils ne participent pas directement aux travaux mais ils président à sa destinée - au CNRS et dans le cadre de l'ATALA - où ils lui accordent un statut disciplinaire, votent les crédits et évaluent les recherches. Outre Martinet et Benveniste, le conseil scientifique du CETA comprend, en 1960, Michel Lejeune, Marcel Cohen, Georges Gougenheim, Bernard Quemada et deux slavistes, Jean Train et Marc Vey. Jean Fourquet les rejoindra plus tard. A l'ATALA, Georges Gougenheim et Marcel Cohen sont des membres particulièrement actifs. André Martinet, Antoine Culioli, Bernard Pottier, David Cohen et Georges Mounin font également partie de l'association.

Dans les sections du CNRS concernées, les linguistes président à l'avenir du nouveau domaine. En 1960, la section 22 « Linguistique générale, langues modernes et littérature comparée », où siègent Benveniste, Martinet et Delavenay, encourage cette « technique encore tâtonnante qui devrait ouvrir des aperçus nouveaux sur bien des faits linguistiques ». Dans le rapport de conjoncture 1961-62, la section 23 « Linguistique française et études littéraires » souligne l'intérêt que présentent pour la linguistique française les recherches entreprises en vue de la TA : « elles obligent à repenser tous les problèmes de linguistique générale, elles offrent aussi l'occasion de prendre mieux conscience des mécanismes fondamentaux de la langue française ». En 1962-63, les deux sections classent la TA au sein d'une nouvelle discipline, la linguistique appliquée, au même titre que l'enseignement des langues avec lequel elle partage l'utilisation de « nouvelles techniques à des fins linguistiques : ordinateur électronique pour la traduction, magnétophone à double piste en pédagogie. »

Cette posture des linguistes, davantage observateurs et législateurs qu'acteurs, va donner une physionomie particulière à la réception des langages formels. Rappelons la situation des linguistes aux Etats-Unis. Sous l'impulsion de Weaver, beaucoup de linguistes, y compris certains distributionnalistes (notamment Z.S. Harris) sont engagés dans des recherches de TA. Certaines caractéristiques du structuralisme américain, analyse distributionnelle, axiomatisation, procédures, les préparent mieux que les Européens à intégrer les méthodes de TA. Ils réfléchissent au problème, « mettent la main à la pâte », ce qui ne les empêche pas d'adopter des attitudes très diverses. Certains pratiquent une linguistique pour la machine qui fait fi des résultats de la linguistique, notamment de la morphophonologie, dans la construction des dictionnaires électroniques (Léon, 1999). C'est l'analyse syntaxique qui fait défaut. Bar-Hillel (1953), dans le sillage de ses travaux sur la syntaxe logique de Carnap, conçoit une syntaxe opérationnelle explicitement élaborée pour la traduction automatique. Apparaissent ainsi les premiers parseurs syntaxiques et les stratégies de parsing, d'abord fondés sur l'analyse en constituants immédiats et la stratégie ascendante (Yngve, 1955), ensuite sur les grammaires syntagmatiques et une stratégie descendante (Yngve, 1960), etc... C'est au travers de ce qui deviendra la linguistique computationnelle (à partir du début des années 1960) que les linguistes-ingénieurs et les linguistes s'approprient la TA, technologie issue des sciences de la guerre et dont la linguistique était *a priori* exclue.

Comme on l'a dit, l'absence de la logique mathématique dans l'horizon de rétrospection des intellectuels français pèse de tout son poids dans cette extériorité des linguistes. Et il faut bien voir que le second tournant de la mathématisation aura besoin là aussi d'intermédiaires, d'opérateurs de passage (Chiss et Puech, 1999), institutions et chercheurs.

3 Opérateurs de passage (1) : Marcel-Paul Schützenberger et Maurice Gross

La création du CETA à l'Institut Blaise Pascal associe donc étroitement la TA au développement du calcul numérique et des calculateurs électroniques. C'est à l'Institut Blaise Pascal que s'élabore un

domaine qui s'apparente beaucoup à ce qui se passe au MIT entre mathématiques, informatique, théorie des automates et langages formels. Deux figures majeures s'y emploient qui joueront le rôle de passeurs entre ce qui se passe aux Etats-Unis et la France. Ce sont véritablement des passeurs et non de simples importateurs, dans la mesure où leurs travaux ont fait avancer les recherches des deux côtés de l'Atlantique et dans les deux sens. L'un comme l'autre ont enrichi les travaux américains comme les travaux français.

Marcel-Paul Schützenberger (1920-1996) peut être considéré comme un véritable passeur entre les mathématiques, l'informatique et les langages formels d'une part, entre les travaux américains et les travaux français d'autre part⁴. D'abord docteur en médecine, puis en mathématiques, ses premiers travaux en informatique théorique sont effectués en France en 1955 et portent sur les codes et la théorie des automates. Dans son premier article, publié dans *l'Evolution psychiatrique* en 1949 « A propos de la cybernétique (Mathématiques et psychologie) », qui constitue une des premières mentions en France de la cybernétique, Schützenberger s'intéresse aux rapports entre la notion d'information développée par Fisher dans les années 1930 et celle de Shannon. Il voit dans la cybernétique et l'information un certain nombre de traits qui auraient dû favoriser leur réception en France par les linguistes structuralistes : tentative d'axiomatisation et organisation abstraite préalable à toute observation (notamment l'utilisation des probabilités *a priori* par Shannon). Comme l'indique Segal (2003 : 427) « C'est la détermination de lois mathématiques régissant les phénomènes avant même leur analyse qui semble apparemment séduire Schützenberger ».

Suite à sa seconde thèse, en mathématiques cette fois, « Contribution aux applications statistiques de la théorie de l'information » soutenue en 1953, il est invité en 1956 par Shannon au MIT, au *Research Laboratory of Electronics* (RLE), pour travailler sur la sémantique des langages formels et, avec Kleene, sur les automates à états-finis. Il y rencontre Chomsky avec lequel, lors d'une seconde invitation aux Etats-Unis en 1963, cette fois à la Harvard Medical School, il élabore le théorème selon lequel tout langage context-free est un codage d'un langage de Dyck, contribuant ainsi à l'édification d'une théorie générale des langages informatiques à l'aide de grammaires context-free. Au milieu des années 1960, il travaille à l'Institut Blaise Pascal au sein d'une équipe chargée de recherches en algèbre appliquée, théorie des langages et des automates comprenant Maurice Gross, André Lentin, Maurice Nivat et Jean-François Perrot. Il se consacre alors essentiellement à l'informatique théorique et fonde la combinatoire des mots, domaine sur lequel il travaille avec André Lentin.

Maurice Gross (1934-2001) peut être considéré comme un second passeur, entre travaux français et américains, entre langages formels et linguistique. Ingénieur de l'armement, il est engagé au CETAP en 1960 pour réaliser un dictionnaire automatique et des programmes d'analyse syntaxique pour la TA. D'octobre 1961 à mai 1962, il effectue un stage de recherche à Harvard sous la direction d'Oettinger qui y dirige le centre de TA, puis au RLE du MIT dirigé alors par Yngve. Il suit les cours de Chomsky au RLE et présente à son séminaire un travail sur la syntaxe de Tesnière. Il discute avec Bar-Hillel des problèmes que pose la TA et de la médiocrité des résultats obtenus. Après la dissolution du CETAP en 1962, il est chercheur CNRS à l'Institut Blaise Pascal et travaille avec Schützenberger sur la théorie des automates et les grammaires formelles. Grâce aux contacts de Schützenberger avec Harris, il est chargé de cours à l'université de Pennsylvanie en 1964-65, et commence des recherches sur sa Grammaire transformationnelle du français selon les méthodes mises au point par Z.S. Harris⁵. De retour à l'Institut Blaise Pascal, Maurice Gross fait partie du groupe « linguistique formalisée » auquel participe également David Cohen qui élabore un analyseur de l'arabe, avant de prendre la succession de Jean-Claude Gardin à la direction de la SAD (Section d'Automatique Documentaire) en 1966 qui deviendra le LADL (Laboratoire d'Automatique Documentaire et Linguistique) en 1970. Jusqu'à sa rencontre avec Harris, Maurice Gross s'intéresse davantage aux problèmes de la toute nouvelle informatique non numérique, orientée vers la théorie des compilateurs et des langages formelles, qu'à la linguistique proprement dite. Après son second séjour aux Etats-Unis et ses travaux menés en collaboration avec Harris, il jouera un rôle de diffusion-information des grammaires formelles au Centre Favard, mais aussi contribuera largement à la diffusion et à l'intégration des travaux harrissiens à la linguistique française.

4 Opérateurs de passage (2) : L'ATALA et le Centre Favard

On a déjà évoqué l'ATALA et son fondateur Emile Delavenay, qui a joué un rôle de passeur-entremetteur déterminant dans le commencement de la TA en France. L'ATALA comprenait des traducteurs, des documentalistes, des ingénieurs, des mathématiciens et des linguistes. Elle a assuré une mission essentielle de diffusion, d'information et d'échanges sur les travaux de TA, puis de linguistique formelle et d'automatisation de la linguistique effectués dans le monde (Léon, 2009).

La formation sur les langages formels et l'informatique est assurée par le Centre Favard et son Séminaire de Linguistique Quantitative créé en mars 1960 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 Jacques Pitrat et Maurice Gross. De cet enseignement, sera issu l'ouvrage de Gross et Lentin *Notions sur les grammaires formelles* paru en 1967, véritable introduction au domaine.

Il est intéressant de noter que ces lieux de diffusion et de formation, bien qu'ouverts à tous, étaient fréquentés par les jeunes chercheurs, mais ne constituaient pas un lieu d'enseignement spécifique pour les étudiants linguistes. Celui-ci n'existait pas à l'université, et la séparation nette entre facultés des sciences et des lettres ne favorisait pas le développement de telles formations où interagissaient plusieurs disciplines. C'est à l'EPHE, 6ème section, qu'une formation de cet ordre a été mise en place pendant quelques années (1967-1970). L'EPRASS (Enseignement préparatoire à la recherche approfondie en Sciences sociales) existait à l'Ecole Pratique des Hautes Etudes depuis 1958 pour former les doctorants africanistes, puis les anthropologues, avec une forte composante de mathématiques et de statistiques dispensée par M. Barbut et Th. Guilbaud. En 1966, elle s'adjoint une section de linguistique et sémantique sous la direction d'A.-J. Greimas qui tente d'organiser un enseignement spécifique pour les linguistes comprenant une formation à la linguistique formelle (logique, formalisation mathématique, programmation, documentation automatique). Y enseignent O. Ducrot, A.-J. Greimas, C. Metz J-B Grize, Y. Gentilhomme, F. François, M. Pêcheux, etc. Cet enseignement, une première en France, a pour objectif de former les linguistes à tous les champs de la linguistique en brisant la tendance existante aux enseignements isolés.

Toutefois les enseignants sont conscients du fait qu'une confrontation trop hâtive entre linguistique et formalisation peut avoir des implications néfastes et contraires aux objectifs souhaités. Deux écueils sont signalés dans la présentation du programme (1966-67). Il est « dangereux pour les chercheurs intéressés par la linguistique générale d'ignorer l'extension donnée par les mathématiques et la logique à la notion de langage, ou les fonctions du langage que révèle son étude psychologique ou sociologique ». Inversement, les étudiants ayant reçu une formation mathématique au départ « considèrent comme allant de soi le découpage de l'énoncé en mots, ou la classification des propositions principales et subordonnées. Il en résulte qu'on applique la logique ou les mathématiques à une conception du langage mise en question depuis longtemps déjà, généralement à celle qui sous-tend les grammaires scolaires d'il y a vingt ans » ; ces différentes postures faisant évidemment obstacle à « toute collaboration ultérieure entre la linguistique et les disciplines formelles ».

Il est mis fin à cet enseignement en 1970. Une lettre de Ducrot justifie l'interruption de la section de linguistique par le fait qu'un enseignement similaire « commence à pénétrer dans les facultés » (lettre de Ducrot 1970, archives EPRASS, EHESS).

5 Les aléas d'une réception externe (1) : la TA comme linguistique computationnelle

Le fait que les travaux en TA ne commencent en France qu'en 1960 a plusieurs conséquences. Alors qu'aux Etats-Unis la linguistique computationnelle s'est construite progressivement à partir de certains projets de TA, en France, TA et linguistique computationnelle sont assimilées et développées simultanément. La phase de TA comme technologie de guerre à peine esquissée par le CETAP (Paris), a

été vite abandonnée dès la dissolution de celui-ci. Le CETAG (Grenoble) a pu travailler d'emblée sur les systèmes de TA en tant que systèmes formels, sans objectif de production de traductions en série.

Le fait d'avoir commencé tard et d'être commandités et soutenus par le CNRS, grand organisme de recherche d'état, a rendu les centres de TA en France moins vulnérables aux critiques des rapports Bar-Hillel et de l'ALPAC, catastrophiques pour la plupart des centres de TA dans le monde. Du moins, les signes négatifs envoyés par le rapport Bar-Hillel - des copies d'une première version du rapport ont circulé en France dès 1959 - a-t-il eu un impact plus diversifié au sein du CETA lui-même. Le CETAP, basé au Fort de Montrouge, doit répondre aux attentes des militaires et devenir un centre de production de traductions d'articles scientifiques et techniques à des fins stratégiques de contre-espionnage. A terme, cette production doit assurer l'autofinancement du centre. Le projet de Sestier de former une équipe d'ingénieurs-linguistes pour développer la TA en tant que technologie linguistique, est vite abandonné devant les maigres résultats obtenus. Sestier démissionne et dissout le CETAP en octobre 1962.

Une fois les militaires désengagés de l'affaire, le CNRS, qui n'est préoccupé que de façon tout à fait secondaire par les aspects économiques de la TA, continue à encourager les travaux du CETA de Grenoble, dirigé par B. Vauquois, pour faire face aux questionnements et aux attentes suscités par les modèles formels et l'analyse syntaxique automatique dans la communauté des mathématiciens appliqués et des linguistes. Il s'agit d'assimiler les langues naturelles à des langages formels et à des langages de programmation et d'explorer, grâce aux systèmes de TA, l'analogie entre traduction et compilation. En 1967, un an après le rapport de l'ALPAC, le CETA n'a jamais été aussi florissant et se voit octroyer un nouveau bâtiment sur le campus de St Martin d'Hères. Il occupe une place internationale en organisant la « Deuxième conférence internationale sur le Traitement Automatique des Langues », la première ayant eu lieu à New York en 1965 sous l'intitulé *International Conference on Computational Linguistics*.

Toutefois le débat des relations entre aspects théoriques de l'automatisation de la linguistique et aspects pratiques de l'ingénierie linguistique n'est pas clos. Aux Etats-Unis, la linguistique computationnelle issue de la TA, malgré tous les efforts de l'ALPAC pour l'ériger en « nouvelle linguistique », continuera à inclure les projets pratiques (y compris la TA) et leurs cortège d'imperfections (linguistique pour la machine incompatible avec les résultats des linguistes, marges d'erreurs etc.). En France, Vauquois est vivement critiqué par Maurice Gross dans son rapport sur le CETA présenté au CNRS en 1967. Pour Gross, qui avait fait partie du CETAP, en avait vécu les échecs, et avait rencontré Bar-Hillel peu de temps après la publication de son rapport, on ne peut pas faire de la linguistique formelle en même temps que de la TA : soit on fait de la TA qui a pour finalité de proposer un produit fini industrialisé et commercial, soit on se consacre à l'analyse des structures des langages artificiels et des langues naturelles. La TA, dit-il, n'est en rien un problème de développement, c'est-à-dire un problème d'adaptation de la linguistique à l'informatique et vice et versa ; c'est un problème pour l'industrie. Durant les décennies suivantes, la tension entre aspects théoriques et aspects appliqués et industriels du TAL demeurera vive (Cori et Léon, 2002).

6 Les aléas d'une réception externe (2) : modèles américains versus modèles soviétiques

En 1960, les Américains n'ont pas seuls l'apanage des modèles de TA. Les Britanniques et les Soviétiques ont aussi développé leurs propres modèles, mais ceux-ci ne sont pas exclusivement fondés sur l'analyse syntaxique. De plus, parmi les modèles d'analyse syntaxique, il existe bien d'autres analyseurs que ceux construits sur les grammaires syntagmatiques. Ainsi les expérimentateurs français vont pouvoir développer des analyseurs en testant et comparant les modèles existants.

Les travaux soviétiques en linguistique mathématique sont bien connus en France. On a vu que les comptes rendus du BSL se faisaient largement l'écho des travaux soviétiques, y compris ceux en linguistique mathématique, davantage même que des travaux américains. Un certain nombre des membres de l'ATALA et/ou du Centre Favard font partie du Parti communiste français. Des chercheurs d'origine roumaine (George Moisil puis Solomon Marcus) assurent la diffusion des travaux des pays du Bloc de

l'Est. L'ATALA reçoit les traductions en français des travaux russes assurées par un centre de l'armée française et est abonnée au centre américain de traductions en anglais des travaux russes (*Joint Publications Research Service*). Contrairement aux expérimentateurs en TA américains qui s'y réfèrent très peu, les membres de l'ATALA publient et discutent les travaux russes dans *TA Informations*, les *Documents de linguistique quantitative* ou la série chez Dunod. C'est le cas des travaux de I.I. Revzin, I.A. Mel'chuk, S.K. Shaumyan, entre autres.

Les travaux des Soviétiques sur la traduction automatique ou la linguistique mathématique de N.D. Andrejev, V.J. Rozencveig, P.S. Kuznecov, A.A. Reformatskij, V.V. Ivanov, et des tchèques Petr Sgall et Lubomir Dolezel sont connus dans le milieu des linguistes en France. L'ouvrage du Roumain Solomon Marcus *Introduction mathématique à la linguistique structurale* qui se donne également pour but d'initier le linguiste et le mathématicien à l'étude de certains aspects mathématiques du langage paraît dans sa version française la même année (1967) que celui de Gross et Lentin, de même que l'ouvrage de Ferenc Kiefer (1968) un an après. L'ouvrage de Marcus fait la part belle aux travaux des pays de l'Est et établit le lien entre les notions mathématiques et les notions de la linguistique structurale de l'Ecole de Prague et des distributionnalistes.

En ce qui concerne le groupe de TA de Grenoble, B. Vauquois et J. Veyrunes se rendent en Union Soviétique en 1962 et prennent contact avec Igor Mel'chuk et le groupe d'Andreev qui travaillent sur des méthodes de langue intermédiaire pour la TA (Archaimbault, Léon, 1997). Vauquois va développer au CETAG son propre modèle de langue intermédiaire, appelée langue-pivot. Proche du modèle de Mel'chuk, c'est un modèle syntaxico-sémantique qui assure l'indépendance des phases d'analyse et de synthèse dans le processus de traduction.

Les modèles français sont aussi mis à contribution et sont comparés avec les modèles chomskyens. Yves Lecerf et David G. Hays voient dans le stemma de Tesnière un type de formalisation. David Hays (1962) élabore à partir de la grammaire de Tesnière son modèle de « *dependency grammar* ». Dans son modèle des conflits, Lecerf (1960) propose de construire une unique représentation en associant automatiquement à une phrase donnée un « graphe de Chomsky » (arbre syntagmatique) et un « graphe de Tesnière » (stemma). La comparaison des deux modèles le conduit à distinguer deux grandes tendances en linguistique : la tendance qui travaille sur les chaînes de mots (Bar-Hillel, Chomsky et Yngve notamment) ; la tendance qui décrit les langues en termes de mots, de « dépendance » ou « hiérarchie » de mots (Tesnière, Hays). Dans le second modèle, on n'a pas besoin de catégories intermédiaires entre la phrase et les mots.

A fin de mettre au point des analyseurs syntaxiques pour la TA, les membres du CETA vont s'appuyer autant sur des modèles américains d'analyseurs, le modèle de « *dependency grammar* » de David Hays ou de « *stratificational grammar* » de Sydney Lamb (1962), que sur des formalismes développés par les Britanniques, comme les structures en treillis (*Cambridge Language Research Unit*) ou les structures arborescentes issues des grammaires chomskyennes. Ce qui va constituer leur originalité, c'est la mise au point d'un langage d'écriture des règles syntaxiques, puis celle du langage-pivot.

7 Les aléas d'une réception externe (3) : les deux types de mathématisation, statistiques et langages formels

7.1 Réception de la théorie de l'information. Deux passeurs

Le courant formel issu des développements en logique mathématique des années 1930-40 et de l'informatisation du langage à partir des années 1950, n'est pas la seule forme de mathématisation du langage de l'après seconde guerre mondiale. Une seconde orientation est issue des travaux statistiques et probabilistes du début du 20^e siècle, qui vont se retrouver remaniés et renforcés par la théorie de l'information. Ceux-ci sont beaucoup plus ancrés dans la tradition française, même s'ils font partie des apports les plus novateurs. Ainsi, dès les années 1950, Marcel Cohen et le groupe qu'il anime, le Centre d'Etudes et de recherches marxistes (CERM) jouent un rôle considérable pour la promotion de la

linguistique statistique, considérée alors comme une des nouvelles approches en linguistique (voir notamment Cohen, 1967).

Les travaux issus des études statistiques et de la théorie de l'information constituent un ensemble très hétérogène de concepts et de méthodes qui vont connaître des destins divers. Suite à la publication de l'ouvrage de Shannon et Weaver en 1949, *The Mathematical Theory of Communication*, la théorie de l'information devient très rapidement ce « bandwagon » dans lequel tout le monde s'engouffre, dénoncé par Shannon (1956), et donner lieu à des travaux disparates.

On peut identifier deux ensembles de travaux et deux passeurs principaux qui vont contribuer à transférer méthodes et concepts de la théorie de l'information en linguistique française. Nous n'évoquerons que pour mémoire les travaux bien connus du premier passeur, Roman Jakobson, qui se donne pour objectif d'appliquer la théorie de l'information à l'analyse des traits distinctifs en phonologie, à partir de l'importation des notions de redondance et du schéma de communication. En France, Martinet d'une part, Dubois de l'autre emprunteront les voies ainsi tracées (Léon, 2008).

Dans une perspective toute autre, et qui va davantage nous intéresser ici, Benoît Mandelbrot (né en 1924) va transférer les concepts de la théorie de l'information vers les études statistiques de vocabulaire, traditionnellement ancrées dans les études stylistiques françaises des années 1940-50, et leur donner un nouvel élan. Mandelbrot est un passeur. Lui aussi a assuré la circulation des théories entre la France et les États-Unis en les enrichissant. Plus encore que Schützenberger et Gross, il poursuit une carrière franco-américaine (impressionnante) dans les laboratoires américains au cœur des sciences de la guerre. Après des études à Polytechnique, il passe deux ans (1947-1949) au Caltech (Institut de technologie de Californie où a enseigné Weaver), puis il travaille un an (1953-54) avec Von Neumann à l'*Institute for Advanced Study* de Princeton. De 1953 à 1971 il enseignera très régulièrement au MIT avant d'enseigner dans d'autres universités américaines et au Collège de France.

Mandelbrot publie en 1954 un article dans la revue *Word* intitulé *Structure formelle des textes et communication*. C'est un travail mathématique sur la loi d'Estoup- Zipf, dont il réexamine les données dans le cadre de la théorie de Shannon. Il s'intéresse aux formes vides (lemmes) par rapport à Zipf qui s'intéresse aux formes fléchies (mots-formes) dont les fréquences auraient des propriétés intrinsèques, ce qui paraît tout à fait discutable à Mandelbrot. Il repense la loi de Zipf comme loi théorique d'organisation des textes. Le débat s'engage avec les stylisticiens qui font des études de fréquence de vocabulaire, comme Pierre Guiraud. Les comptes rendus d'ouvrage et les articles sur les statistiques lexicales se multiplient et se croisent dans le BSL des années 1950 – comptes rendus de Mandelbrot par Guiraud, et de Guiraud par Mandelbrot, compte rendu de Herdan par Mandelbrot - de même que les exposés au sein de la SLP elle-même. Ce débat fait donc partie intégralement de la linguistique en France.

Il est intéressant de noter, qu'au-delà de la linguistique, la théorie de l'information est très diffusée en France dès les années 1950. La cybernétique et la théorie de la communication intéressaient les psychologues et les philosophes. Ainsi, le *Journal de Psychologie normale et pathologique* se fait l'écho de ces nouvelles théories par les nombreux comptes rendus parus dès 1955 d'ouvrages en anglais ou en français sur le domaine, notamment les ouvrages de Grey Walter, Ross Ashby, G.-A. Miller. Par ailleurs, les traitements statistiques, au cœur de certains champs de la psychologie (psychologie différentielle notamment) bénéficient d'une attention particulière. Ce qui contribue sans doute à expliquer la faveur spéciale faite à Pierre Guiraud dont les ouvrages sont systématiquement recensés⁶.

La question reste de déterminer s'il s'agit véritablement d'une mathématisation de la linguistique et si elle a apporté des résultats. Pêcheux (1969) considère que ces méthodes mettent en oeuvre un concept pré-saussurien (et donc pré-scientifique), à savoir la bi-univocité du rapport signifiant-signifié. Auroux (2009) qualifie de « mathématisation couvrante » les approches quantitatives qui abordent le langage « par-dessus », en effectuant des comptages d'éléments observables, comme les mots, et en leur attribuant une propriété ou un chiffre ; il l'oppose à la « mathématisation substitutive », celle opérée par les langages formels, qui substituent au langage des variables et des constantes.

Ce débat, qui n'a cessé d'exister en France, prendra un nouveau tournant dans les années 1990 avec la mise à disposition des grands corpus informatisés, qui débute avec les débuts de la lexicométrie, les travaux de Charles Muller, de Bernard Quemada et la mise en place du TLF au début des années 1960. Dans les années 1960, l'automatisation de ce courant concurrence le courant formel en France.

7.2 Guerre des dénominations et frontières de champs

Une des conséquences de l'existence du champ spécifique des études statistiques, considéré comme négligeable par les chomskyens et la linguistique computationnelle, mais profondément ancré dans la tradition française et ravivé par la théorie de l'information, est la tension provoquée par la rivalité entre les deux orientations. En témoigne la multitude des dénominations qui vont désigner le champ et les sous-champs du domaine, que depuis le début des années 1960 les Américains désignent par '*computational linguistics*'.

Le Centre Favard opte explicitement pour l'appellation '*séminaire de linguistique quantitative*' afin de ne pas dissocier les aspects formels de la linguistique et les méthodes statistiques.

Desclés et Fuchs (1969) discutent une classification proposée par Solomon Marcus. Pour ne reprendre que quelques classes, Marcus distingue la linguistique algébrique (travaux sur les monoïdes comme ceux de Chomsky-Schützenberger) ; la linguistique mathématique (utilisant les chaînes de Markov) dont la linguistique probabiliste et la linguistique quantitative ; la linguistique automatique, computationnelle ou cybernétique ; la linguistique appliquée. Marcus n'inclut pas dans la linguistique computationnelle les travaux sur les grammaires formelles ni les statistiques de vocabulaire. C'est le terme de linguistique formelle qui subsume l'ensemble.

Vauquois en 1969, alors président de l'ATALA considère que la *computational linguistics* recouvre deux types de recherche :

- (i) Les travaux de linguistes conduits par des linguistes et qui utilisent des ordinateurs à titre d'outil : analyse de concordances, calculs de fréquence de mots, recherches lexicographiques, confection de dictionnaire à consultation automatique, et surtout grammaires génératives.
- (ii) Les équipes pluridisciplinaires formées de linguistes et de mathématiciens dont l'objectif n'appartient exclusivement à aucune de ces disciplines, mais au traitement automatique des langues proprement dit. Un tel projet recouvre une masse d'applications, comme la communication homme-machine, l'enseignement programmé, la documentation automatique, les études de formalisation sémantique, enfin la traduction automatique. (Vauquois, 1969 :60)

Ce qui est remarquable, et très probablement plus lié à des raisons de politique scientifique qu'à des considérations épistémologiques, c'est de regrouper grammaire générative et études statistiques de vocabulaire d'un côté et traitement automatique des langues et formalisation sémantique de l'autre⁷. Par ailleurs, ces classifications ne font pas apparaître les distinctions opérées par la « Computational Linguistics ». Celle-ci déclare englober, tout en les distinguant explicitement, les aspects théoriques de l'interaction langages formels, linguistique, programmation d'une part, et les aspects pratiques de l'ingénierie linguistique d'autre part. Ceci aura des conséquences sur la constitution du TAL dans les années ultérieures.

Cette profusion montre la diversité des champs concernés mais aussi la complexité de leur intégration dans le champ des sciences du langage en France. Là où la première mathématisation avait fait défaut, la seconde automatisation-mathématisation se trouve confrontée à des sources particulièrement disparates (automatisées ou non, américaines ou soviétiques, grammaires formelles ou études statistiques de vocabulaire, théoriques ou orientées vers les applications à visée industrielles) la plupart du temps complètement externes à la tradition intellectuelle en France et à l'horizon de rétrospection de la linguistique française.

8 Les aléas d'une réception externe (4) : réception et/ou réflexivité, l'externalité redoublée

Sur le plan de l'historiographie et des études réflexives sur la mathématisation du langage, la situation en France est aussi très différente de celle des Etats-Unis. Jusqu'au milieu des années 1960, c'est Chomsky lui-même qui assure la synthèse des travaux sur les grammaires formelles (voir entre autres Chomsky et Miller, 1963). Le relais est pris dans les années 1980, une fois les limitations formelles du modèle transformationnel reconnues, par des ouvrages réhabilitant les modèles hors contextes face aux grammaires transformationnelles (voir entre autres Savitch et al., 1987). Plus généralement, c'est dans la revue *Computational Linguistics* que se trouvent les articles critiques sur la formalisation de la linguistique. Les langages formels et les modèles de mathématisation du langage sont la plupart du temps associés à l'algorithmique, aux langages de programmation et plus généralement à l'informatique (Perrault, 1984).

En France, faute d'enracinement propre et d'intérêt précoce pour le domaine, les premiers travaux sur la formalisation s'avèrent d'emblée en partie réflexifs. On a vu que les recherches sur l'analyse syntaxique se faisaient à partir de la comparaison des modèles existants. Parallèlement, faute de cursus spécifiques, se manifeste immédiatement une demande de cours et de manuels d'introduction au domaine. Aussi ne s'étonnera-t-on pas que l'ouvrage d'initiation aux langages formels de Gross et Lentin (1967) paraisse une dizaine d'années avant son homologue américain *Fundamentals of Mathematics for Linguistics* publié en 1978 par Barbara Hall Partee. Le premier est issu des séminaires de linguistique quantitative du Centre Favard, destinés à la formation ou plus simplement l'information des linguistes, alors que le second est un cours donné par l'auteur à partir 1962, dans le cadre d'un cursus universitaire, et qui évolue chaque année.

Dès la fin des années 1960, en complément de la revue de l'ATALA, intitulée alors *TA Informations*, la revue *Mathématiques et Sciences humaines*, avec des numéros spéciaux consacrés au thème « Mathématiques et linguistique », font régulièrement le point sur le domaine (en 1971, 1982 et 1988). Ils comparent les formalismes grammaticaux, relativisent l'apport de la grammaire générative, puis dans les années 1980 notent les évolutions comme le retour de la logique, l'émergence de l'intelligence artificielle et d'autres modèles de mathématisation, théorie des graphes ou topologie...

L'externalité se voit assumée en quelque sorte par la réflexivité. Mais la question reste, même si des formes d'institutionnalisation ont pris le relais comme par exemple des cursus spécifiques de TAL, de savoir si l'intégration de l'automatisation-mathématisation dans les sciences du langage en France a réellement réussi⁸.

9 1965-66 : « nouvelle génération » de linguistes et « nouvelle linguistique »

Un mot, avant de conclure, sur le début de l'intégration par le milieu linguistique de la linguistique formelle. On peut situer celle-ci à partir des années 1965-66.

Au début des années 1960, les rares jeunes linguistes qui lisent les travaux américains le font de façon isolée et en fonction de leur trajet personnel. Les témoignages recueillis par Chevalier (2006) manifestent de cette singularité. Dans le cadre de sa thèse soutenue en 1955, Pottier lit tous les travaux américains qui lui sont accessibles, dans *Word* et *Language*. Mais il n'est pas un diffuseur, il se sert éventuellement des analyses pour son propre travail mais n'en fait pas matière à enseignement⁹. Gross et Ruwet lisent *Syntactic Structures* au tout début des années 1960, l'un en 1961 avant son premier séjour aux USA, le second en 1960 après l'avoir trouvé sur le bureau de Lacan. L'un et l'autre avouent n'avoir compris le texte qu'après s'être familiarisé avec les travaux américains, en particulier avec ceux de Harris ; lecture systématique pour Ruwet, collaboration étroite dans le cas de Gross. Ce sont en effet avant tout les textes de Harris qui leur ont permis de pénétrer dans le contexte linguistique américain, que leur formation et environnement français (ou belge) rendaient inaccessible.

Ce n'est qu'en 1965-66, que les travaux américains, parmi lesquels les travaux sur les grammaires formelles, vont être diffusés de façon systématique auprès des linguistes voire d'un public plus large, au-delà du cercle de l'ATALA et du Centre Favard. C'est aussi en 1966 que Gross qui, selon son propre témoignage, se sentait jusqu'alors avant tout informaticien, isolé des linguistes qu'il ne fréquentait qu'au Centre Favard, commence à s'insérer dans le milieu linguistique français.

La politique de *Langages* (créée en 1966), voulue par Jean Dubois, de traduction systématique et de diffusion des travaux américains est un vecteur important de l'intégration de la linguistique formelle, comme en témoignent un certain nombre des premiers numéros dirigés par Todorov, Ducrot, Ruwet, Gross ou Dubois.

Le séminaire de Culioli de 1963 est transformé en laboratoire de linguistique formelle et les séminaires de l'AFLA, en particulier les deux premiers à Besançon en 1965, et à Nancy en 1966 se consacrent à la « nouvelle linguistique »¹⁰.

10 Conclusion

Pour conclure, un mot sur la méthode. Afin d'appréhender cette étude sur la réception de l'automatisation-mathématisation, qui comporte un émetteur (ou plutôt un site d'origine) autant qu'un récepteur, nous avons opté pour une étude historique institutionnelle et comparative. Toutefois notre focale est restée centrée sur la réception, donc sur le domaine français. Une étude institutionnelle et comparative permet de définir les contextes historiques et scientifiques, les ancrages, à partir desquels va s'opérer la réception par rapport aux contextes d'origine. Le processus de réception externe exige avant tout de déterminer les opérateurs de passage. Les sociétés savantes ont joué le rôle de diffusion et d'« entremetteurs » entre les chercheurs de disciplines qui auparavant ne se fréquentaient pas, mathématiciens et linguistes principalement. Il a fallu également déterminer des passeurs, chercheurs d'exception comme M.-P. Schützenberger et B. Mandelbrot dont le rôle n'est pas simplement d'importer des concepts et des méthodes, mais de transformer les théories dans le site de réception, et aussi dans le site d'origine. En l'absence de configuration des sciences de la guerre, totalement inédite et limitée aux Etats-Unis, des institutions et des disciplines ont dû prendre le relais pour initier les recherches. Commencer à faire de la TA n'était pas une mince affaire. En France, c'est le CNRS avec l'Institut Blaise Pascal et les mathématiques appliquées qui ont joué le rôle de catalyseur.

D'autres tournants de la mathématisation ont eu lieu depuis. Pour les étudier, il faudra changer de focale. Le tournant des études sur corpus des années 1990, par exemple, exigerait de mettre en avant les travaux britanniques et ceci dans une mise en perspective historique qui excède largement les cinquante dernières années, et ne s'appréhende que très partiellement par une histoire institutionnelle.

Références bibliographiques

Archives du CNRS : Fonds documentaire du CNRS de Gif-sur-Yvette.

Archives de l'ATALA.

Documents de l'EPRASS (archives personnelles de S.Fisher et archives de l'EHESS).

Bulletin de la Société Linguistique de Paris (1942-1970).

Archambault, S., Léon, J. (1997). La langue intermédiaire dans la Traduction Automatique en URSS (1954-1960). Filiations et modèles. *Histoire Epistémologie Langage*, 19-2, 105-132.

Auroux, S. (1987). Histoire des sciences et entropie des systèmes scientifiques. Les horizons de retrospection. In Schmitter P. (éc.) *Zur Theorie und Methode der Geschichtsschreibung der Linguistik*, Tübingen : Narr., 20-42.

Auroux, S. (2006). Les modes d'historicisation. *Histoire Epistémologie Langage*, 28-1, 105-116.

Auroux, S. (2009). Mathématisation de la linguistique et nature du langage. *Histoire Epistémologie Langage*, 31-1, 5-45.

- Bar-Hillel, Y. (1953). Some linguistic problems connected with Machine Translation. *Philosophy of Science*, 20, 217-225.
- Cantineau, J. (1954). Compte rendu de Harris Z.S., 1951, Methods in Structural Linguistics. *BSL* 50-2 : 4-9.
- 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* 63, 57-102.
- Chevalier, J-C. (1990). La linguistique au CNRS 1939-1949. *Cahiers pour l'histoire du CNRS*, 1990-9, 39-80.
- Chevalier, J-C. (2006). *Combats pour la linguistique, de Martinet à Kristeva*. Lyon : ENS Editions.
- Chiss, J-L. & Puech, C. (1999). *Le langage et ses disciplines, 19^e 20^e siècles*, Paris & Bruxelles : Editions Duculot.
- Chomsky, N. & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In Luce, D., Bush, R. & Galanter, E. (éds.), *Handbook of Mathematical Psychology Vol. II*, New York: Wiley, 269-321.
- Cohen, M. (1967). Sur l'histoire de la statistique en linguistique. *Etudes de linguistique appliquée* 5, 3-8.
- Cori, M. & Léon, J. (2002). La constitution du TAL. Etude historique des dénominations et des concepts. *Traitement Automatique des Langues*, 43-3, 21-55.
- Dahan A. & Pestre, D. (eds.). (2004). *Les sciences pour la guerre (1940-1960)*. Paris : Editions de l'EHESS.
- Desclés, J-P. & Fuchs, C. (1969). Le séminaire international de linguistique formelle. *TA Informations*, 1969-1, 1-5.
- Dosse, F. (1991). *Histoire du Structuralisme. I. Le champ du signe 1945-1966*, Paris : Ed. La Découverte.
- Guiraud, P. (1959). Compte rendu de William N .Locke et A. Donald Booth Machine Translation of language 1955. *Journal de Psychologie normale et pathologique*, 52-4 : 483.
- Hays, D. G. (1962). Automatic language-data processing. *Computer Applications in the Behavioral Sciences*, H. Borko ed. Prentice Hall, Englewood Cliffs, N.J., 395-421.
- Kiefer, F. (1968). *Mathematical linguistics in Eastern Europe*, New York : Elsevier.
- Lamb, S.M. (1962). On the mechanization of syntactic analysis. *Proceedings of the International Conference on Machine Translation and Applied Language Analysis, Teddington 1961* London: HMSO, 673-686. (traduction française : 1964, « La mécanisation de l'analyse syntaxique », Traduction automatique et linguistique appliquée. Choix de communications présentées à la Conférence internationale sur la Traduction mécanique et l'Analyse linguistique appliquée, National Physical Laboratory Teddington 1961, ATALA (ed.), Paris : PUF, 169-183.
- Lecerf, Y. (1960). Programme des conflits, modèle des conflits. *La Traduction automatique*, 4-5, 17-36.
- Léon, J. (1998). Les débuts de la traduction automatique en France (1959-1968) : à contretemps ? *Modèles Linguistiques*, 19- 2, 55-86.
- Léon, J. (1999). La mécanisation du dictionnaire dans les premières expériences de traduction automatique (1948-1960). In Cram D., Linn A. & E. Nowak (eds.). *History of Linguistics 1996 vol. II*, John Benjamins Publishing Company, 331-340.
- Léon, J. (2008). Théorie de l'information, information et linguistes français dans les années 1960. Un exemple de transfert entre mathématiques et sciences du langage. In Durand J., Habert B., Laks B. (éds.), Actes du Congrès Mondial de Linguistique Française , Paris, 9-12 juillet 2008 : 923-938.
- Léon, J. (2009). Atala : cinquante ans. <http://www.atala.org/ATALA-cinquante-ans>.
- Locke, W.N. & Booth A.D. (eds.). (1955). Machine Translation of Languages, 14 Essays. New York : MIT and John Wiley & son.
- Mandelbrot, B. (1954). Structure formelle des textes et communication. *Word* 10-3, 1-27.
- Marcus, S. (1967). *Introduction mathématique à la linguistique structurale*. Paris : Dunod.
- Murray, S. (1994). Theory groups and the study of language in North America. John Benjamins Publishing company.
- Pêcheux, M. (1969). *Analyse automatique du discours*. Paris : Dunod.

- Perrault, R. C. (ed.) (1984). Special Issue on Mathematical Properties of Grammatical Formalisms. *Computational Linguistics* 10/3-4.
- Ramunni, J. (1989). *La physique du calcul, Histoire de l'ordinateur*. Paris : Hachette.
- Savitch, W. J., Bach E., Marsh W., & Safran Naveh G. (eds.). (1987). *The Formal complexity of Natural Language*. Dordrecht : D. Reidel.
- Shannon, Cl. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana : University of Illinois Press.
- Shannon, C. E. (1956). The Bandwagon, Institute of Radio Engineers, Transactions on Information Theory, Vol. IT-2, 3.
- Vauquois, B. (1969). Dix ans d'ATALA: de la traduction automatique au traitement automatique des langues. *TA Informations*, 1969-2, 57-61.
- Wagner, R-L. (1954). Compte rendu de Harris Z.S., 1951, Methods in Structural Linguistics. *Journal de psychologie normale et pathologique* 47-4 : 537-539.
- Yngve, V.H. (1955). Syntax and the problem of multiple meaning. In Locke W.N. & Booth A.D. (eds.). *Machine Translation of Languages, 14 essays*, New York : MIT and John Wiley & son, 208-226.
- Yngve, V.H. (1960). A Model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104-5, 444-466.

¹ Pour les deux mouvements de mathématisation, on parlera de mathématisation intrinsèque (Auroux 2009) dans la mesure où ils affectent tous les deux les concepts en linguistique.

² Sur les débuts de la traduction automatique en France voir Léon (1998).

³ Il est intéressant de noter que le CNRS fondé en 1939 et destiné à regrouper les « sciences humaines et sociales » et les sciences appliquées sous le contrôle de comités spécialisés composés de savants, d'industriels, d'ingénieurs, comporte une organisation comparable à celle du MIT aux Etats-Unis, à savoir un centre de recherche où interagissent sciences, sciences de l'ingénieur et sciences humaines. Sa mise en sommeil pendant la guerre a probablement affaibli l'originalité de cet objectif premier.

⁴ Segal (2003), qui lui consacre une étude approfondie dans son histoire de la théorie de l'information, souligne le caractère original, voire unique des travaux de Schützenberger, d'ailleurs trop peu reconnus en France.

⁵ La réception des travaux de Z.S. Harris est un exemple tout à fait particulier de la réception de la linguistique américaine et même de la formalisation en France, et demanderait à être approfondie.

⁶ Il faut noter que le *Journal de Psychologie normale et pathologique* et le *Bulletin de la SLP* entretiennent des liens étroits. Le *Journal de Psychologie* publie régulièrement des articles (notamment de Benveniste ou de Martinet) et recense les ouvrages de linguistes. Quant au *BSL*, il consacre des comptes rendus aux numéros du *Journal de Psychologie* traitant du langage.

⁷ En fait, ce que traduit la classification de Vauquois de même que l'apparition du terme de « linguistique formelle » comme générique de l'ensemble dans celle de Marcus, c'est l'émergence du courant culiolien de formalisation sémantique à la fin des années 1960. Son examen excède l'étude de la réception que nous menons ici.

⁸ Le TAL est devenu un thème à part entière de CML10 alors qu'il ne l'était pas en 2008.

⁹ On peut toutefois avancer que cette connaissance approfondie des travaux américains a joué un rôle non négligeable dans le fait que Pottier ait été le seul linguiste à entreprendre des travaux pour la TA au début des années 1960.

¹⁰ « Hippolyte, Aigrain et Nivat m'ont demandé de faire un cours de linguistique qui pourrait intéresser les mathématiciens. J'ai décidé d'appeler ce cours Séminaire de Linguistique formelle. J'ignorais totalement qu'il existait des *Papers in Formal Linguistics* de Harris et j'ai inventé le terme de linguistique formelle ou plutôt je me le suis inventé. Je lisais énormément de choses, mais pas Harris » (A. Culioli dans Chevalier et Encrevé, 1984 :92).