

Linguistique appliquée

et traitement automatique des langues

Étude historique et comparative

Jacqueline Léon

UMR7597, Cnrs, Université Paris Diderot, Université Paris Sorbonne Nouvelle, Sorbonne Paris Cité

Résumé

Cet article examine les relations entre la linguistique appliquée et le traitement automatique des langues au début de leur institutionnalisation dans les années 1960. Dans une étude comparative, on essaiera de rendre compte des différences entre traditions. En particulier on essaiera d'expliquer pourquoi des liens forts ont existé en France, notamment au sein de l'association Atala, ainsi qu'au niveau international au sein de l'Aila, alors que dans les autres pays pionniers, comme la Grande-Bretagne et les États-Unis, la LA s'est davantage orientée vers d'autres centres d'intérêt comme l'instrumentation de l'enseignement des langues et l'analyse statistique des textes. Enfin, on montrera comment une réflexion approfondie en Urss sur les applications de la linguistique a nourri l'interaction entre LA et TAL.

Mots-Clés

linguistique appliquée, traitement automatique des langues, traduction automatique, histoire

Abstract

This article examines the relationship between applied linguistics and natural language processing at the beginning of their institutionalization in the 1960s. In a comparative study, we attempt to account for differences between traditions. In particular we will try to explain why strong links existed in France, particularly in the Atala association, as well as internationally in the Aila, while, in other pioneering countries such as Britain and the United States, applied linguistics turned towards other areas of interest such as instrumentation language teaching and the statistical analysis of texts. Finally, we show how in-depth reflection on linguistic applications in the Ussr fed the interaction between LA and TAL.

Keywords

applied linguistics, natural language processing, machine translation, history

1. Introduction

Alors que, l'Aila 2014 a proposé, pour commémorer ses cinquante ans, de revisiter les thèmes du colloque fondateur de 1964, on s'aperçoit que le traitement automatique des langues (désormais TAL), en l'occurrence la traduction automatique, y a occupé la première place. Depuis, les relations entre le TAL et la linguistique appliquée (désormais LA) se sont distendues et ne sont plus qu'épisodiques. Une étude historique des liens entre les deux

domaines permet de montrer le rôle majeur qu'ils ont joué dans les processus d'institutionnalisation et dans l'autonomisation de la linguistique appliquée en France. Notre étude portera sur la période 1950-1979 et nous interrogerons les points suivants : A quoi correspond cette association entre LA et TAL au sein du 1^{er} colloque de l'Aila de 1964 ? Est-elle seulement conjoncturelle ou bien est-elle le fruit d'une réflexion sur le rapport entre automatiser le langage et l'enseignement des langues ? Y a-t-il eu pendant cette période une attention particulière portée à l'ensemble des applications de la linguistique ou bien plus spécifiquement aux applications de l'informatique à la linguistique ? L'appartenance du TAL à la LA est-elle une spécificité française ? Qu'en est-il des autres traditions, Grande-Bretagne, USA et URSS ?

2. La situation française : l'Atala et l'Aila

L'association entre traduction automatique (désormais TA) et LA est à relier tout d'abord au rôle de Bernard Pottier (né en 1924). On le sait, l'Aila a été créée à l'issue d'un colloque qui a eu lieu à Nancy en 1964. Ce sont Bernard Pottier et Guy Bourquin (né en 1928), pionniers de la traduction automatique en France, qui l'organisent¹. Ils créent un centre de TA à la Faculté des Lettres et Sciences Humaines de Nancy en 1962 ainsi qu'un Centre de Recherches de Linguistique Appliquée. Le Certificat de Linguistique Appliquée de 1962-63 comporte à l'oral une option 'traduction automatique'. Bernard Pottier est le premier président de l'Aila mais il est aussi un des membres fondateurs de l'Atala (Association pour l'étude et le développement de la Traduction Automatique et de la Linguistique Appliquée) dont il sera un des vice-présidents de 1966 à 1971 et le président de 1971 à 1974. Parmi ses thèmes d'études et dans son nom même, l'Atala fait coexister traduction automatique et linguistique appliquée, sans lien d'inclusion ni de préséance.

¹ Pottier et Bourquin sont les seuls linguistes français qui ont développé un système de TA. Les autres pionniers sont des mathématiciens et des ingénieurs.

En France, la linguistique appliquée est ancrée dans les études de vocabulaire et la lexicologie, domaines de prédilection des linguistes français dans les années 1940-50. Bernard Pottier a d'ailleurs effectué ses premières expériences de traduction automatique en sémantique lexicale, considérant que le lexique est un des domaines les plus complexes de la TA, tant au plan formel (notamment les mots composés) que sur le plan sémantique (polysémie).

Il est intéressant de voir la façon dont les linguistes des sections 22 et 23 du Cnrs considèrent la linguistique appliquée à l'aube de son institutionnalisation. Pour la section 22 'Linguistique générale, langues modernes et littérature comparée', qui comprenait, entre autres, les linguistes Émile Benveniste et André Martinet et un physicien, Bernard Vauquois (1929-1985), directeur du Centre d'Etude en Traduction Automatique de Grenoble (Cetag), le terme 'linguistique appliquée' englobe plusieurs choses indiquant son caractère composite (Candel, 2011). Dans le rapport de conjoncture du Cnrs de 1963-64 (section 22), la linguistique appliquée désigne a) l'enseignement des langues b) les applications informatiques de la linguistique (TA, documentation automatique etc.).

L'appellation de 'linguistique appliquée' recouvre deux ordres d'application qui posent chacun des problèmes spécifiques: a) la pédagogie de l'enseignement des langues... Il y aurait lieu de songer ici à un institut de linguistique appliquée, comme il en existe en Grande-Bretagne et aux Etats-Unis. b) la traduction automatique, pour laquelle un centre de recherches fonctionne déjà, et à laquelle des liens étroits associent la documentation automatique (la langue documentaire posant elle aussi des problèmes de structuration sémantique) qui intéresse de nombreuses disciplines (rapport de conjoncture du Cnrs 1963-64, section 22 : 94).

Pour la section 23 'linguistique française et études littéraires', ce qu'elle appelle linguistique appliquée, qui inclut certes la TA, reste vague. Ce qui l'intéresse, ce sont les recherches sur la langue française :

On voudrait souligner l'intérêt que présentent pour la linguistique française les recherches entreprises en vue de la TA : outre qu'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. De manière plus générale, on assiste au développement d'une discipline nouvelle ou renouvelée, la linguistique appliquée (rapport de conjoncture du Cnrs de 1963-64, section 23 : 103)².

Même dans les années pionnières de la TA en France, l'idée d'une association entre LA et TA n'est pas unanimement partagée, ni par les acteurs de la TA ni par ceux de la LA. Ainsi, le fondateur de l'Atala, Émile Delavenay (1905-2003), ne mentionne pas du tout la LA dans son *Que-sais-je ? La traduction automatique* publié en 1959 pas plus que dans la bibliographie sur la TA qu'il publie avec son épouse Katharine Delavenay en mars 1959. Le premier colloque européen de TA qui a lieu à Teddington en 1961 s'intitule 'First International Conference on Machine Translation of Languages and Applied Language Analysis' où la TA est associée à l'analyse appliquée au langage ; il s'agit de l'analyse automatique appliquée au langage qui ne concerne en rien l'enseignement des langues. De façon parallèle, il est à noter que le Centre de Linguistique Appliquée de Besançon créé en 1958 par Bernard Quemada (né en 1926), ne comporte pas de section de TA, et que la revue *Études de Linguistique Appliquée* créée en 1962, ne publiera ni articles de TA ni articles de TAL mais s'intéressera aux travaux en statistique linguistique.

Avec l'institutionnalisation de la linguistique appliquée (Linn et *al.*, 2011) et les déboires de la TA au milieu des années 1960, la séparation entre LA et TAL va s'accroître. A sa création, l'Aila est l'acronyme de 'Association Internationale de Linguistique Appliquée à

² En 1969, toute mention du lien entre TA et linguistique appliquée a disparu du rapport de conjoncture du Cnrs. La TA est déconsidérée et a laissé la place au TAL alors en voie de constitution. Dans la partie du rapport consacrée au TAL, dont Bernard Vauquois, membre de la section 'Linguistique générale, langues et littératures étrangères' – devenue la section 28 -, est le probable rédacteur, toute référence à la LA a disparu.

l'enseignement des langues vivantes' et non de 'Association Internationale de Linguistique Appliquée' tout court, qu'elle deviendra ensuite. Autrement dit, à ses débuts, la 'linguistique appliquée' inclut plusieurs domaines, dont l'enseignement des langues. En 1965, la revue de l'Atala *La Traduction Automatique* change de nom suite au discrédit porté à la TA par le rapport Bar-Hillel publié en 1960 et la constitution du comité en 1964 - dont le rapport est publié en 1966 - qui va mettre fin aux financements de la TA aux USA et dans le monde. Elle prend le nom de *TA Informations. Revue internationale des applications de l'automatique au langage* que Vauquois, devenu président de l'Atala en 1966, glose de la façon suivante :

Un tel projet recouvre alors une masse d'applications dont la traduction automatique n'est qu'un exemple. La communication (et même la conversation homme-machine) est une application fort intéressante qui aujourd'hui connaît une vague d'enthousiasme. Des sujets tels que l'enseignement programmé et la documentation automatique vont de toute évidence recevoir un éclairage nouveau à mesure que ces modèles se seront développés. Enfin, les études de formalisation sémantique que ces modèles exigent sont en général des sujets hors des frontières linguistiques ou mathématiques (Vauquois, 1969 : 61).

Cette autonomisation du TAL se concrétise également dans le fait que la revue *TA Informations* accueille dès 1965 le bulletin de l'Aila nouvellement créée, sous forme d'une rubrique au titre bilingue 'Nouvelles de l' Aila. *Applied Linguistics News*'³. La création d'un bulletin autonome pour l'association de linguistique appliquée accompagne l'institutionnalisation des deux domaines qui vont continuer cependant à s'entraider sous la forme d'une alliance objective. Comme l'indique Delavenay, président fondateur de l'Atala, dans son texte de présentation de la refonte de la revue, où est également signalé

³ Cet hébergement durera de 1965 à 1970. Par ailleurs le premier bulletin de l'Aila ne paraît que beaucoup plus tard en automne 1978 (Cori, Léon, 2002).

l'hébergement du Bulletin de l'Aila, l'autonomisation de deux domaines est nécessaire. Comme on le voit également dans ce texte, le traitement automatique du langage a désormais remplacé la TA. La vocation de l'Atala est de fournir un état des lieux

des nombreux centres où les méthodes nées du calcul automatique sont appliquées à la nouvelle linguistique. Les divers domaines de la linguistique appliquée seront couverts, avec une préférence pour le traitement automatique du langage, mais sans exclusive; il faudra encore quelque temps avant que ne se différencient nettement les disciplines diverses de la linguistique appliquée, et que chacune trace un sillon profond et reconnaissable dans la direction qui lui est propre. Jusqu'à ce jour, et jusqu'à ce que chacune dispose de ses propres moyens et méthodes d'information, la vocation de cette série restera partiellement au moins interdisciplinaire, comme en témoigne la participation rédactionnelle de l'AILA (Delavenay, 1965 : 2).

L'Afla, créée en 1965, reste toutefois nettement impliquée dans le TAL. Elle organise son deuxième séminaire de linguistique générale et appliquée au Ceta (centre de Traduction Automatique) de Grenoble et invite son directeur Bernard Vauquois à faire un cours sur 'La théorie formelle des langages et la traduction automatique'. Ce cours, selon les auteurs du compte-rendu, constitue un des 'clous' du séminaire, cours associé à un atelier de 'travaux mécanolinguistiques' où les participants ont pu s'initier au langage *Comit*⁴ (TA Informations, 1967-2 : 97-105). Le compte-rendu conclut en affirmant que la LA est une science jeune avec de multiples problèmes concrets à aborder : pathologie du langage, traitement automatique des informations de type linguistique, enseignement des langues.

Ainsi malgré l'institutionnalisation et les tentatives d'autonomisation des deux domaines, ils restent indissociablement liés et cela pendant de nombreuses années. C'est ainsi que le

⁴ *Comit* est le premier langage de programmation capable de traiter des chaînes de caractère, mis au point en 1955 pour la traduction automatique par Victor Yngve qui dirigeait le groupe de traduction automatique au *Research Laboratory of Electronics* du MIT.

rapport du Cnrs sur les Applications des Mathématiques aux Sciences de la Société et à la Linguistique, publié en 1984, inclut le TAL au sein de la LA entendue comme ensemble des applications de la linguistique :

On doit s'accorder à reconnaître, à côté de la linguistique pure, l'existence d'une linguistique appliquée (LA) avec en note : le domaine de la LA comprend essentiellement deux grandes parties : l'une vouée au traitement automatique des langues et du langage (donc en gros la linguistique computationnelle), l'autre aux applications de la linguistique à l'enseignement des langues. Ces deux parties ont en commun ce qui touche à l'EAO (Enseignement Assisté par Ordinateur) pour les langues (Rapport Cnrs, 1984 : 33).

L'enjeu est donc de faire exister et de légitimer, à côté de la linguistique pure, des applications de la linguistique. Le premier numéro du bulletin de l'Afla (paru en 1978) tente de définir ce qu'est une application en linguistique, sans toutefois engager une réflexion sur le plan théorique, pourtant devenue nécessaire :

L'adjectif 'appliquée' accolé au nom 'linguistique' dans l'expression 'linguistique appliquée' signifie que ce qui est l'objet de la recherche, ce n'est pas la linguistique en tant qu'étude du fonctionnement de la langue, mais la linguistique en tant qu'elle est indissolublement liée à l'accomplissement d'un certain nombre de tâches dont l'objectif est autre que la seule étude du fonctionnement de la langue, mais qui ne peuvent se faire indépendamment de la linguistique [...]. [Il y aurait] nécessité d'une réflexion sur une théorie de l'application, mais ce n'est pas le but ici (Bulletin de l'Afla, 1978 : 1).

Pour conclure ce point, disons qu'en France l'alliance TA et enseignement des langues au sein de la linguistique appliquée a contribué à l'institutionnalisation du TAL et à celle de la LA. Avec la création de l'Aila, les deux domaines vont s'autonomiser. Mais cela va s'effectuer de façon asymétrique : l'Aila (et à suite l'Afla) va continuer à considérer le TAL et les applications informatiques à la linguistique comme faisant partie de son champ, alors que l'Atala, qui héberge le bulletin de l'Aila à ses débuts, ne considèrera plus l'application de la linguistique comme un de ses domaines de recherche.

3. Linguistique appliquée et TAL dans la tradition britannique : de la traduction automatique à la Corpus Linguistics

L'institutionnalisation de la linguistique appliquée en Grande-Bretagne commence avec la création, à l'initiative de John C. Catford (1917-2009), de l'École de linguistique appliquée à l'Université d'Édimbourg en 1957, consacrée à la formation des enseignants en langues étrangères. La création de l'association britannique de linguistique appliquée, Baal, *British Association for Applied Linguistics*, sous la présidence de Pit Corder (1918-1990), suit de quelques années celle de l'Aila et de l'Afla. Celle-ci a lieu en 1967, dans le sillage de la nomination de Peter Strevens (1922-1989) à la chaire de linguistique appliquée de la toute nouvelle Université d'Essex fondée en 1964.

Contrairement à l'Aila dont les thèmes sont à la fois la traduction automatique et l'enseignement des langues, Baal se consacre essentiellement à l'enseignement des langues, à l'étude de l'usage et de l'acquisition du langage. Ainsi que l'indique la présidente de Baal, Rosamund Mitchell, lors de la célébration des trente ans de l'association en 1997 :

There was some initial debate over the scope of the Association's objectives, and in particular, whether it should concern itself primarily/ exclusively with matters to do with language teaching, and machine translation (then the stated objectives of Aila,

which were to be imitated). At the Reading meeting, however, a wider brief was agreed, accepting that the Association could legitimately concern itself with applications of linguistics much more broadly. In the 1974 version of the Constitution (formalised at that time to meet Charity Commissioners' requirements), the aims of Baal were finalised to read:

'The Objects of the Association are the advancement of education by fostering and promoting, by any lawful charitable means, the study of language use, language acquisition and language teaching, and the fostering of interdisciplinary collaboration in this study...' (Mitchell, 1997 : 5).

De même, le second colloque de l'Aila organisé à Cambridge en 1969, comporte très peu de communications en traitement automatique des langues. À l'exception de la conférence invitée d'un pionnier de la TA et co-auteur du rapport de l'Alpac, David G. Hays (1928-1995), sur le thème '*applied computational linguistics*', il n'y a qu'une seule session (section 2) en traitement automatique, comportant trois communications (Perren & Trim, 1971). Son intitulé '*computer analysis of text*' montre l'intérêt des Britanniques pour l'analyse automatique des textes préfigurant leur rôle essentiel dans l'apparition de la *Corpus Linguistics* (Léon, 2008).

La notion d'application est fortement ancrée dans la conception britannique des sciences du langage (Léon, 2011). Contrairement à la France, les applications sont inscrites, depuis le XIX^e siècle, dans le programme même de la linguistique générale tout en étant source d'innovations technologiques. Henry Sweet (1845-1912) conçoit une nouvelle notation phonétique, le *Broad Romic* - sur laquelle sera fondé l'IPA-, une méthode de sténographie et une réforme de l'orthographe. John Ruppert Firth (1895-1960) élabore, en collaboration avec Daniel Jones (1887-1967), un projet d'orthographe mondiale (*a practical script*) fondée sur l'alphabet latin et destinée à l'imprimerie, à l'enseignement, et aussi au développement de la linguistique où il constitue un cadre pour une notation scientifique.

Firth développe de nouveaux instruments pour la phonétique et préconise le recours aux langages restreints pour la traduction et pour établir les grammaires des différentes variétés d'anglais et les grammaires scolaires. Halliday et *al.* (1964) inclut la traduction automatique comme domaine d'application de la linguistique et utilise les registres, version remaniée des langages restreints de Firth, pour promouvoir une conception 'pratique' de l'enseignement des langues.

En Grande-Bretagne comme en France, on trouve des acteurs engagés à la fois dans la TA et dans la linguistique appliquée. C'est le cas de MAK Halliday (né en 1925), éminent pionnier de la TA au sein du *Cambridge Language Research Unit* et auteur d'un imposant article dans le premier numéro d'*ELA* en 1962 intitulé 'Linguistique générale et linguistique appliquée'⁵.

Halliday est un des rares Britanniques à défendre la TA comme partie intégrante de la linguistique appliquée⁶. La TA est un problème d'application de la linguistique, seule la linguistique permet de formuler les règles qui seront introduites dans le programme. Cette application de la linguistique met en jeu des spécialistes de différentes disciplines, dont des mathématiciens et des ingénieurs. La linguistique appliquée ne peut se limiter à l'enseignement des langues, mais inclut des domaines beaucoup plus vastes comme la TA et la socio-linguistique, de même que des domaines n'appartenant pas à la linguistique mais essentiel pour l'enseignement des langues comme la psychologie et les sciences de l'éducation :

⁵ Il est intéressant de noter que la version anglaise de l'article reproduit dans McIntosh & Halliday (1966 : 1-41) a pour titre *General Linguistics and its application to language teaching*, qui met davantage l'accent sur l'enseignement des langues.

⁶ On notera d'ailleurs que Halliday n'a jamais fait partie des dirigeants de la Baal.

Machine translation is a problem of applied linguistics. It is not of course a question of programming a computer to perform operations in linguistic theory, nor even just of inserting a linguistic description into the store, although the latter is in one sense part of the total operation. What does go into the programme, however, can be found out only by linguistic methods (Halliday et al., 1964: 132).

One of the principal applications of linguistics today is in machine translation ; this involves many people besides linguists, but they are all specialists of some kind, mathematicians, computer programmers and others (ibid. : 139).

This term ['applied linguistics'] covers the whole field of research and training in the teaching of languages, including but not restricted to general linguistics and general phonetics. It is misleading because it excludes other activities which are just as much applications of linguistics, such as machine translation and sociological linguistics. At the same time, it includes parts of certain subjects which lie outside the linguistic sciences, such as those aspects of psychology and educational theory that are relevant to language teaching (ibid. : 166).

Halliday est le seul à s'intéresser à la LA parmi les pionniers de la TA en Grande-Bretagne ; ses travaux ultérieurs en traitement automatique des langues, se situent dans la tradition britannique, notamment firthienne, d'études des collocations dans les textes. Il a supervisé le projet Osti (*UK Government Office for Scientific and Technical Information*) entrepris en 1963 par John Sinclair, alors jeune chercheur à l'Université d'Édimbourg. Ce projet est destiné à l'étude par ordinateur des *patterns* de collocations à partir d'enregistrements d'anglais oral et écrit. Ce sont les premiers travaux en *Corpus Linguistics* préfigurant ceux consacrés aux grands corpus informatisés dont il a été un des premiers promoteurs (Léon, 2008).

L'Aila est une création européenne, et même si les orientations françaises sont dominantes, le bureau de l'Aila comprend, outre son président Bernard Pottier, trois secrétaires, le Britannique Peter Strevens, la Belge Lydia Hirschberg, et le Suédois Max Gorosch de Stockholm, qui vont infléchir ses positions sur la LA.

Dans sa présentation de l'Aila rédigée en anglais, Peter Strevens, dont on a vu l'importance dans l'institutionnalisation de la LA en Grande-Bretagne, distingue deux aspects : la linguistique appliquée à l'enseignement et les traitements statistiques linguistiques à grande échelle (*applied linguistics in large-scale statistical treatment of language*). Considérant que le traitement automatique des langues est déjà largement représenté dans les institutions, ce sont les traitements statistiques que l'on doit privilégier. Avec cette recommandation, Strevens rejoint à la fois les positions de Halliday et Sinclair, et celles de Bernard Quemada à Besançon, sur l'importance de l'analyse statistique des textes, qui, dans les années 1990, une fois les masses de données linguistiques rendues disponibles par les développements technologiques des ordinateurs et des logiciels, aboutira à la *Corpus Linguistics*.

Lydia Hirschberg est une mathématicienne. Elle est chargée du bulletin de l'Aila (qui a pour nom 'Newsletter' dans *TA Informations*). Elle fonde en 1964 le *Centre de linguistique automatique appliquée* à Bruxelles. Dans ce centre, est élaboré Dicautom (Dictionnaire automatique de terminologie destiné à aider les traducteurs et les interprètes en langages techniques), à la demande de la Ceca (Communauté européenne du charbon et de l'acier)⁷.

Le projet de *linguistique automatique appliquée* implique d'étudier les limites de la formalisation du langage naturel à partir de données linguistiques expérimentales mais sans simplifier les hypothèses comme le fait la linguistique mathématique qui ne se fonde sur aucune donnée linguistique empirique.

⁷ Un colloque de linguistique appliquée organisé en 1966 à Luxembourg par la Ceca et l'Aila, à destination des traducteurs et interprètes, fut entièrement consacré à la terminologie et la lexicographie. 'Les journées ont culminé en une démonstration du Dicautom...' (Kukenheim, 1967 : 27).

Autrement dit, au sein de l'Aila sont préconisées d'autres applications automatisées du langage et pas seulement la traduction automatique et l'enseignement des langues. Ces applications, traitements statistiques sur corpus et dictionnaires automatisés, sont reconnues de plein droit comme appartenant à la linguistique appliquée.

4. Linguistique appliquée et TA aux États-Unis. Effort de guerre et instrumentation

Aux États-Unis⁸, on trouve la même dissymétrie qu'en Grande-Bretagne (et qu'en France à partir de 1965), à savoir que les centres de TA ne s'occupent pas d'enseignement des langues alors que les centres de LA s'intéressent aux applications de l'informatique. Pour l'Amtcl (*Association for Machine Translation and Computational Linguistics*) créée en 1962, ce qu'on appelle linguistique appliquée c'est la linguistique computationnelle définie comme l'application des ordinateurs à la linguistique. En revanche, la principale revue de LA, les *Monograph series on languages and linguistics*, traite régulièrement de la TA dès les années 1950. Dans le sillage de l'énorme effort fourni pendant la guerre 1939-45, dans le domaine de l'enseignement des langues, la LA se donne deux objectifs principaux : mener des recherches sur l'enseignement de l'anglais langue étrangère et tester de nouvelles méthodes et du nouveau matériel pour cet enseignement, magnétophones, méthodes audiovisuelles, laboratoires de langue. Cet objectif est clairement affiché dans la revue créée en 1948 par Charles Fries (1887-1967), le *Quarterly Journal of Applied Linguistics* (où le terme *applied linguistics* apparaît pour la première fois). C'est aussi celui du *Centre for Applied Linguistics* créé en 1959 sous la direction de Charles A. Ferguson (1921-1998) par la *Modern Language Association*, et dont la mission est principalement de faire face à la pénurie de personnel et de matériel face à la demande d'enseignement de l'anglais à l'étranger, la nécessité d'étudier les langues d'Asie et de former les enseignants de langue en

⁸ Sur l'histoire des relations entre linguistique appliquée et traduction automatique aux États-Unis, voir Léon (2015).

linguistique. Quant au traitement automatique des langues, il ne fait pas partie de ses missions :

The Center for Applied Linguistics is expected to play a crucial role in this situation in the first place by assessing the current demand and available supply of personnel and materials, by exploring the status of research projects on the analysis of languages (*in particular the major languages of the Middle East and South and Southeast Asia*), and by assembling and distributing materials and information. Important operations of the Center will include the regular publication of a newsletter, the setting up of language and country files, and the amassing of a library of existing pedagogical materials (The Linguistic Reporter, éditorial du premier numéro, 1959: 1).

De fait, très peu de linguistes se sont trouvés à la croisée de l'enseignement des langues et des technologies de guerre, même si celles-ci auront des retombées importantes sur la technologisation de l'enseignement. Deux personnalités sont ici à signaler, Martin Joos et Léon Dostert.

Les linguistes qui ont travaillé au chiffre à Arlington Hall (Virginie), siège de l'*US Army's Signal Intelligence Service*, sont pour la plupart employés comme traducteurs pour les langues rares. Martin Joos (1907-1978), avec une double formation d'ingénieur et de phonéticien, est le seul qui ait participé à la conception et au développement de méthodes et d'appareillage servant au codage et décodage. En collaboration avec les Laboratoires Bell (où il côtoie probablement Claude Shannon), il travaille à l'amélioration du *Sound Spectrograph*, instrument le plus en pointe à l'époque en matière d'enregistrement et d'analyse du son. Les spectrographes et la phonétique acoustique ont joué un rôle central dans l'élaboration des technologies pour l'enseignement des langues, l'analyse et la synthèse de la parole, et plus généralement le développement des laboratoires de langue. Joos écrit un

manuel d'enseignement du néerlandais et, en 1945, conçoit un appareil, le *Speech Stretcher*, destiné à aider les étudiants à améliorer leur prononciation.

Léon Dostert (1904-1971) fait également figure d'exception en alliant des activités d'enseignement des langues, de traduction et de traduction automatique. Dostert a travaillé comme interprète lors du procès de Nuremberg. Il fonde l'*Institute of Languages and Linguistics* à l'Université Georgetown en 1949 et organise tous les ans des tables rondes sur les rapports entre linguistique et enseignement des langues qui sont publiées dans les *Monograph series on languages and linguistics*. Il est nommé Président de la *National Federation of Modern Language Teachers Associations* en 1960. Par ailleurs, Dostert joue un rôle fondamental dans les débuts de la TA. Il dirige le projet de Georgetown qui aboutit à la création de Systran, un système de TA existant encore à l'heure actuelle sous une forme interactive sur internet. En janvier 1954, il organise à New York la première démonstration de TA sur ordinateur, en collaboration avec IBM. Comme Joos, Dostert invente un instrument pour l'enseignement des langues, un magnétophone double piste grâce auquel l'étudiant peut s'enregistrer et se réécouter juste après avoir entendu le modèle. Il n'est donc pas surprenant que l'*Institute of Languages and Linguistics* ait été un pionnier de l'utilisation de technologies de pointe pour l'enseignement. La table ronde de 1957 est entièrement consacrée à la TA et, dès 1955, la revue *Monograph series* publie régulièrement des articles dans le domaine.

Ainsi, aux États-Unis, ce lien étroit entre développements technologiques, enseignement des langues et linguistique appliquée doit être interprété dans le cadre de la culture de guerre. Le contexte est très différent du contexte européen où l'Aila voit le jour dans le cadre du développement des communautés européennes, notamment de la Ceca. Cet ancrage spécifique de la linguistique appliquée aux USA peut expliquer que les Américains ne se sont doté que tardivement d'une association de linguistique appliquée. Ils ont déjà de nombreuses institutions créées dans l'immédiat après guerre et consacrées à la LA, parmi lesquelles le *Centre for Applied Linguistics* et l'*Institute of Languages and Linguistics*. Il est probable que L'Aaal (*American Association for Applied Linguistics*) a été créée en 1977 dans le but spécifique de rallier l'Aila.

5. En Urss une réflexion théorique sur les applications de la linguistique

Face à la dissymétrie observée entre TA et LA chez la plupart des pionniers, il reste à expliquer cette particularité française, à savoir que TA et enseignement des langues se sont trouvés associés au moment de la création de l'Atala, premier centre de TA français.

Pour répondre à cette question, il faut examiner la situation en Urss, en rappelant que les Français, pour des raisons diverses, connaissent très bien les travaux soviétiques, contrairement aux Britanniques ou aux Américains qui n'en font quasiment jamais mention (voir Léon, 2010). Beaucoup de chercheurs en linguistique ou en TA font partie du Parti communiste français. Les comptes-rendus du *Bulletin de la Société de Linguistique* se sont fait largement l'écho des travaux soviétiques, dont ceux en linguistique mathématique, davantage même que des travaux américains. L'Atala reçoit les traductions en français des travaux russes assurées par le service de recherche de l'armée française (le Casden, Comité d'action scientifique de défense du centre d'exploitation scientifique et technique du ministère de la Défense) 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 malgré les traductions, les membres de l'Atala publient et discutent les travaux russes. En ce qui concerne le groupe de TA de Grenoble, Bernard Vauquois et Jean Veyrunes se rendent en Union Soviétique en 1962 et prennent contact avec Igor A. Mel'čuk (né en 1932) de l'Institut de Linguistique de l'Académie des Sciences de Moscou) et le groupe de Nikolaj D. Andreev (1920-1997), de l'Université de Leningrad) qui travaillent sur des méthodes de langue intermédiaire pour la TA. Vauquois développe au Cetag son propre modèle de langue intermédiaire, appelée langue-pivot, proche du modèle de Mel'čuk.

En Urss, la TA et la LA sont associées dans les revues, les colloques et les centres de TA. Victor Ju. Rozencvejk (1911-1998) renomme en 1959 sa revue consacrée à la TA 'revue de TA et de linguistique appliquée' *Mashinnyy Perevod i Prikladnaya lingvistika*, publiée au

sein d'un institut pédagogique (le Premier Institut Pédagogique d'Etat des Langues étrangères de Moscou) où sont menées des recherches théoriques sur les problèmes de TA et de linguistique appliquée (notamment la formalisation de la syntaxe). Il existe une section de LA, *The structural and applied linguistics Department*, dirigée par Alexandre A. Reformatskij au sein de l'Institut de Linguistique de l'Académie des Sciences où travaille Mel'čuk. Les colloques de Linguistique appliquée de Cernovtsky (Ukraine) en 1960 et en 1961 comportent des communications en méthodologie pour l'enseignement des langues, en linguistique mathématique, en traduction et en TA (Papp 1966). C'est donc en Urss qu'on trouve ce lien entre enseignement des langues et TA.

Par ailleurs il existe une véritable réflexion sur les sciences appliquées. L'article d'Andreev et Zinder publié en 1959 dans *Voprosy Yazykoznanja* intitulé *Basic problems in Applied Linguistics* définissent la LA comme un ensemble de pratiques linguistiques, qui disent-ils, ont souvent été ignorées au profit de la linguistique théorique. Ils inscrivent ces préoccupations dans la tradition russe en alléguant notamment que la notion de LA a été introduite par Baudouin de Courtenay à la fin du XIX^e siècle.

Loin de se limiter à l'enseignement des langues et à la traduction automatique, les auteurs montrent que les pratiques linguistiques sont très variées et doivent toutes se fonder sur des savoirs linguistiques. Ils hiérarchisent l'urgence de la mise en œuvre de ces pratiques en fonction des besoins propres de l'Union Soviétique. Ainsi, la multiplicité des langues nécessite que l'on s'occupe de façon prioritaire des systèmes d'écriture pour les langues non écrites, des systèmes de transcription et de translittération pour la cartographie et pour harmoniser les noms propres dans les catalogues de bibliothèques et les bibliographies scientifiques. La diffusion des sciences implique que l'on s'intéresse à la terminologie en participant aux commissions internationales de standardisation. La sténotypie et les problèmes liés à l'orthoépie et à la logopédie ont été inclus depuis longtemps dans le champ de la LA. La traduction scientifique et technique doit s'appuyer sur des règles linguistiques, et doit faire partie de la LA. Ils insistent aussi sur l'importance de la linguistique pour les technologies de la communication, dans le but d'établir des statistiques de vocabulaire et des constantes de la parole (paramètres acoustiques de la parole) pour lesquelles ils préconisent

l'effort conjoint d'ingénieurs, physiologistes, psychologues et linguistes (en se référant notamment aux travaux de Roman Jakobson, Gunnar Fant et Morris Halle). Enfin ils militent pour un code universel de la science, considérant que la correspondance entre ce code et les diverses langues nationales est un des problèmes majeurs de la linguistique appliquée et exige la participation des linguistes.

En ce qui concerne les relations entre linguistique et enseignement des langues, ils rappellent que Lev Šcerba (1880-1944) est un des rares linguistes qui s'intéressaient à l'enseignement des langues. Selon lui, il s'agit plus d'une discipline linguistique que pédagogique, elle doit être une branche de la linguistique générale et être fondée sur une théorie générale du bilinguisme.

Concernant plus spécifiquement la TA, ils signalent, de façon tout à fait intéressante, que les aspects linguistiques des applications reposaient jusqu'à présent essentiellement sur le caractère acoustique du langage parlé⁹ alors que cet aspect est négligeable pour la traduction automatique de textes écrits de même que pour la recherche d'informations automatisée. Ils faut donc s'attacher prioritairement à la morphologie, à la syntaxe et la sémantique pour la traduction automatique.

Enfin, Andreev et Zinder sont sans doute les seuls à mentionner l'aide que la TA peut apporter à l'enseignement des langues : ainsi des parties d'algorithmes morphologiques et syntaxiques élaborés pour la TA ont été appliquées avec succès à l'enseignement des langues.

Les auteurs insistent sur l'importance de la théorie linguistique pour toutes ces pratiques, et déplorent qu'il n'existe pas encore de théorie générale de l'écriture ni de théorie générale de la traduction. Ils soulignent également l'importance des mathématiques. Les nouveaux problèmes de la linguistique appliquée ne peuvent pas être résolus par des méthodes traditionnelles. En plus d'aspects qualitatifs, ces problèmes ont des aspects quantitatifs qui exigent l'utilisation de méthodes mathématiques. Est préconisée la formation mathématique

⁹ Voir les travaux de Joos sur les spectrographes mentionnés ici même § 4, et ceux de Jakobson Fant et Halle sur la théorie des traits distinctifs, discutés par Andreev et Zinder.

des linguistes de sorte qu'un domaine émerge, celui de la 'linguistique et mathématiques appliquées'.

Les pionniers français de la TA connaissent très bien ces réflexions sur les applications de la linguistique et des mathématiques. René Moreau, mathématicien spécialiste de la théorie de l'information, cite Andreev et Zinder (1959) dans ses séminaires. Mounin (1964), citant Delavenay, signale que les Soviétiques ont fondé une linguistique nouvelle, la *prikladnaja lingvistika*, ensemble d'applications de la linguistique :

C'est [...] à une conception nouvelle des études de linguistique que nous invite la recherche de l'automatique en traduction [...] qui nécessiterait 'la création, au sein des disciplines linguistiques de nos universités, d'une sous-discipline nouvelle orientée vers des buts pratiques, vers une technologie de l'avenir' [Delavenay, la machine à traduire p.13] c'est elle qui a abouti à la prikladnaja lingvistika, née semble-t-il (comme sous-discipline) chez les Russes, et passée déjà en plusieurs endroits dans les structures universitaires sous forme d'Institute for Applied Linguistics (Washington) ou de Cours de Linguistique Appliquée (Paris) (Mounin, 1964 : 54).

6. Conclusion

Pour conclure, on peut avancer que l'association entre TA et LA était au départ de nature conjoncturelle dans la situation française. Elle correspondait à une concomitance particulière des processus d'institutionnalisation des deux domaines au début des années 1960. Cette association s'est distendue après la création de l'Aila et la dissymétrie de traitement des deux domaines l'un par l'autre s'est retrouvée identique à celle observée dans les autres

traditions, notamment britannique et américaine. La LA accueille le TAL parmi ses thèmes de recherches, alors que le TAL ne compte plus ni la linguistique appliquée ni l'enseignement des langues parmi ses thématiques. Seule la technologisation de l'enseignement des langues peut être considérée comme un champ commun aux deux domaines.

Toutefois la situation française s'explique aussi par l'influence des travaux soviétiques, notamment l'ampleur du périmètre de la LA en Urss, définie comme un ensemble de pratiques linguistiques incluant les applications des mathématiques ou de l'informatique à la linguistique, associé à une réflexion théorique approfondie. Cette richesse des travaux russes, bien connus des chercheurs français, contribue à expliquer l'association de l'enseignement des langues et de la TA au sein de la LA en France jusqu'aux années 1980.

Dans la période actuelle, la mise à distance du TAL à l'égard de la LA participe d'un mouvement plus général d'éloignement du TAL à l'égard de la linguistique. La prédominance des méthodes statistiques en TAL enlèvent leurs prérogatives aux domaines de la linguistique traditionnellement sollicités pour l'élaboration de règles (syntaxe, morphologie ou sémantique). La linguistique est au mieux instrumentalisée et les structures linguistiques quand on les utilise encore sont réduites à de simples données, au même titre que les nombres, les tableaux, et les 'sacs de mots'¹⁰. Le TAL ne s'intéresse plus à la linguistique appliquée, parce qu'il ne se compte plus parmi les applications de la linguistique.

Références

Archives du Cnrs. Fonds documentaire du Cnrs de Gif-sur-Yvette.

¹⁰ Voir sur ce point le numéro spécial de *l'Information grammaticale* n°142 consacré aux traitements automatiques de l'oral et de l'écrit et à la place de la linguistique.

Archives *Histoire du Traitement Automatique des Langues* Htal (UMR7597, Histoire des Théories Linguistiques).

Bulletin de l'Afla, n°1 automne 1978.

TA. Informations 1965-1969, revue de l'Atala.

The Linguistic Reporter. (1959). vol.1-1. Bulletin du *Center for Applied Linguistics*.

Andreev, N.D. & Zinder, L.R. (1959). 'Basic problems in Applied Linguistics'. *Voprosy Yazykoznanja* n°4, pp. 3-19 [Tr. Angl. JPRS 991-D, 29 oct 1959 pp. 18-40].

Bar-Hillel Y. (1960). 'The present Status of Automatic Translation of Languages'. In F.C. Alt (dir.) *Advances in Computers* vol. 1, N.Y., London : Academic Press, pp. 91-141.

Candel, D. (2011). 'Linguistique appliquée : parcours définitoires et lexicographiques'. *Histoire Epistémologie Langage*, vol. 33, n°1, pp. 99-115.

Cori, M. & Léon, J. (2002). 'La constitution du TAL. Étude historique des dénominations et des concepts'. *Traitement Automatique des Langues*, vol. 43, n°3, pp. 21-55.

Delavenay, E. (1965). 'Perspectives de TA Informations'. *TA Informations*, n° 1-2, pp.1-3.

Halliday, M.A.K. (1962). 'Linguistique générale et linguistique appliquée'. *Études de Linguistique Appliquée* n°1, pp.5-42. [version anglaise 'General Linguistics and its Application to Language Teaching' In McIntosh & Halliday (1966). pp. 1-41, et Halliday M.A.K. (2007). pp.135-173].

Halliday, M.A.K. (2007). 'Applied Linguistics as an Evolving Theme', In J. Webster (dir.) *Collected Works of MAK Halliday Language and Education*, vol 9, London, New York : Continuum, pp.1-16.

Halliday, M.A.K., McIntosh A. & Stevens P. (1964). *The Linguistic Sciences and Language Teaching*, London : Longmans, Green and co ltd.

Kukenheim, L. (1967). 'Compte-rendu du colloque de linguistique appliquée'. *Nouvelles de l'Aila, TA Informations*, n°1, pp. 27-33.

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.

Léon, J. (2008). 'Aux sources de la *Corpus Linguistics* : Firth et la London School'. *Langages*, n°171, pp. 12-33.

Léon, J. (2010). 'Automatisation-mathématisation de la linguistique en France dans les années 1960. Un cas de réception externe'. In F.Neuve, V.Muni-Toke, J.Durand, T.Kingler, L.Mondada & S.Prévoist (dir.), *Actes du 2e Congrès Mondial de Linguistique Française*, pp. 825-838. Paris : EDP Sciences. <http://www.linguistiquefrancaise.org>[DOI10.1051/cmlf2010158]

Léon, J. (2015). *Histoire de l'automatisation des sciences du langage*, Lyon : ENS Editions.

Léon, J., Luzzati, D. & Poibeau, T. (dir.) (2014). 'Traitements automatiques de l'oral et de l'écrit (2). Quelle place pour la linguistique ?', *L'information grammaticale* n°142.

Linn, A., Candel, D. & Léon, J. (dir.) (2011). 'Disciplinarisation de la linguistique appliquée', *Histoire Epistémologie Langage*, vol. 33, n°1.

McIntosh, A. & Halliday, M.A.K. (1966). *Patterns of Language. Papers in General, Descriptive and Applied Linguistics*, London : Longmans.

Mitchell, R. (1997). *Notes on the history of the British Association for Applied Linguistics (1967-1997)*. Baal. <http://www.baal.org.uk/>

Mounin, G. (1964). *La machine à traduire. Histoire des problèmes linguistiques*. La Haye : Mouton.

Papp, F. (1966). *Mathematical Linguistics in the Soviet Union*. London, The Hague, Paris : Mouton & co.

Perren G. E. & L. Trim (1971). *Applications of Linguistics, Selected Papers of the Second International Congress of Applied linguistics, Cambridge 1969*. Cambridge : CUP.

‘Rapport du Cnrs sur les Applications des Mathématiques aux Sciences de la Société et à la Linguistique’, Linguistique et disciplines intéressées par les langues et le langage, *Mathématiques et Sciences Humaines* n°86, pp.31-45, 1984.

Vauquois B. (1969). ‘Dix ans d’Atala : de la traduction automatique au traitement automatique des langues’. *TA Informations* n°2, pp. 57-61.

À propos de l’auteure

Jacqueline Léon est directrice de recherche émérite au Cnrs, au laboratoire d’Histoire des Théorie Linguistiques (HTL), université Paris Diderot, université Sorbonne Nouvelle, Sorbonne Paris Cité. Ses recherches portent sur l’histoire de l’automatisation des sciences du langage, l’histoire du structuralisme, l’histoire de la linguistique britannique et de la linguistique américaine.

Courriel : jacqueline.leon@univ-paris-diderot.fr

Toile : <http://htl.linguist.univ-paris-diderot.fr/>

Adresse : UMR7597 HTL, Université Paris Diderot, UFR Linguistique, case 7034, 5 rue Thomas Mann, 75205 Paris cedex 13, France.