

VARIA

**LA LANGUE INTERMÉDIAIRE DANS LA TRADUCTION
AUTOMATIQUE EN URSS (1954-1960)
Filiations et modèles ¹**

Sylvie ARCHAIMBAULT, Jacqueline LÉON
URA CNRS 381 - Université Paris 7

¹. Nous tenons à remercier Sylvain Auroux, Irina Ivanova, André Lentin et Patrice Pognan pour leur lecture attentive et leurs précieuses remarques et surtout Igor Mel'uk pour les informations nombreuses et inédites qu'il nous a si généreusement fournies. Ils ne sauraient bien sûr être tenus responsables des éventuelles erreurs qui subsisteraient dans ce texte.

RESUME : Les premières recherches en *Traduction Automatique* (TA) débutèrent en URSS en 1954. Elles se distinguent des recherches américaines, commencées en 1948, par leur caractère théorique et le choix d'une méthode faisant appel à une langue intermédiaire. La création de langues intermédiaires prend sa source dans la linguistique comparative de Baudouin de Courtenay et dans la réflexion sur l'internationalisation des langues. Le travail pionnier de P.P. Trojanskij de 1933 est réactualisé par Panov en 1956, qui conçoit la langue intermédiaire comme devant être une langue naturelle, et spécifiquement le russe. Se développèrent parallèlement des modèles sémantiques de langue intermédiaire comme ceux d'Andreev ou de Mel'uk, modèles qu'on peut apparenter à des tentatives de langues universelles.

ABSTRACT : Research work on *Automatic Translation* (TA) started in the USSR in 1954. It differed from American research, started in 1948, by its theoretical slant and by the choice of a method that resorted to an intermediate language. The concept of intermediate languages springs from Baudouin de Courtenay's comparative linguistics and from a reflection on the internationalisation of languages. In 1956, P. P. Trojanskij's pioneering work of 1933 was updated by Panov, who sees the intermediate language as a natural language, and more specifically Russian. At the same time, semantic models for intermediate languages were developed by, for instance, Andreev or Mel'uk. These models may be seen as attempts to construct a universal language.

MOTS-CLES : Histoire de la linguistique ; URSS ; Traduction Automatique ; Langue intermédiaire ; Langue universelle ; Sémantique ; XX^e siècle ; Baudouin de Courtenay ; Drezen E.K. ; Trojanskij P.P. ; Panov D. Ju. ; Andreev N.D. ; Mel'uk I.A.

KEY WORDS : History of linguistics ; USSR ; Machine translation ; Intermediary language ; Universal language ; Semantics ; 20th century ; Baudouin de Courtenay ; Drezen E.K. ; Trojanskij P.P. ; Panov D. Ju. ; Andreev N.D. ; Mel'uk I.A.

INTRODUCTION

LA RECHERCHE en traduction automatique (désormais TA)² commença en URSS en 1954 à la suite de la première démonstration sur calculatrice électronique qui eut lieu en janvier 1954 à New York. D. Ju. Panov eut ainsi connaissance des travaux du groupe de l'Université de Georgetown, notamment du programme de TA élaboré par le linguiste Paul Garvin et l'ingénieur Paul Sheridan. Toutefois, les Soviétiques prirent d'emblée des options différentes pour des raisons à la fois théoriques, technologiques et pratiques.

Comme les groupes américains, les groupes soviétiques comprennent peu de linguistes. Mais ces linguistes se rattachent à une solide tradition de lin-

². La présente étude s'appuie principalement sur des documents publiés. Nous avons fait appel autant que possible, pour les compléter, à des témoignages d'acteurs contemporains. On sait toutefois que, dans la Russie soviétique, l'activité scientifique était très fortement encadrée et contrôlée par le parti communiste, et que carrières scientifiques et politiques étaient souvent étroitement mêlées. En conséquence, on ne peut être sûr que les travaux les plus intéressants aient tous été publiés. On peut craindre d'ailleurs que cet état de chose ne perdure aujourd'hui (*cf.* Cejtin 1997).

guistique comparative³ qui les amènera à poursuivre une réflexion théorique approfondie sur le passage d'une langue à l'autre et sur le processus de traduction multilingue. Par ailleurs, les groupes soviétiques comprennent un certain nombre de mathématiciens qui, face à la pénurie de machines électroniques, presque entièrement monopolisées par les militaires et les services secrets, ont développé des algorithmes de préférence à des systèmes de TA "complets", c'est-à-dire comprenant un module programmé. Cette situation a une conséquence importante : les chercheurs sont moins soumis à la pression sociale et aux contraintes économiques ou politiques leur imposant de fournir des résultats rapides et performants. Les recherches peuvent donc être résolument théoriques. En outre, les chercheurs en TA, engagés dans des disciplines comme les mathématiques ou la linguistique mathématique, considérées à cette période comme peu "dangereuses" idéologiquement, bénéficiaient d'une certaine tranquillité, qui, même si elle les marginalisait en partie, leur ont permis de jeter les bases d'une linguistique originale, de telle sorte que le mathématicien V.A. Uspenskij (né en 1937), qui a participé à ces travaux, a qualifié l'époque 1955-1967 de "cercle d'argent" de la linguistique russe. Enfin, les Soviétiques se trouvent confrontés à des problèmes linguistiques et algorithmiques spécifiques liés à la direction dans laquelle s'opère la traduction. La traduction anglais/français vers le russe, à laquelle ils doivent s'atteler, impose des contraintes très différentes de la traduction du russe comme langue source, qui est le lot des Américains.

La commande des autorités en URSS est spécifique. Outre les besoins de traduction scientifique et technique, préoccupation commune à tous les états au lendemain de la Seconde Guerre mondiale, il existe en URSS une volonté politique de traduction du russe en diverses langues de l'Union. La traduction multilingue correspond à une nécessité.

Si, dès le départ, les Soviétiques confrontent leurs propres recherches aux travaux américains, du moins à ceux qui leur sont accessibles, ils cherchent à développer des conceptions et des méthodes originales et enracinées dans une continuité scientifique et culturelle. La langue intermédiaire constitue précisément l'un des thèmes les plus originaux des recherches russes en TA.

La question de la traduction multilingue et l'idée d'utiliser une langue intermédiaire n'étaient certes pas totalement absentes des premiers travaux américains, ainsi qu'en témoignent plusieurs articles du premier recueil de TA, publié en 1955 par A.D. Booth et W.N. Locke. Reifler préconisait d'utiliser, à des fins de traduction multilingue, une langue "à grammaire et morphologie simple" telle que le chinois comme *pivot language*. Booth évoquait la création d'une langue artificielle, *le Machineese* (y avait-t-il jeu de mots ?), susceptible

³. Cette tradition est entretenue en Russie à travers l'influence de Baudouin de Courtenay (1845-1929) et de F.F. Fortunatov (1848-1914), fondateur de l'école linguistique de Moscou. Elle imprimera sa marque propre au développement de tous les domaines de la linguistique et reste très vivace dans la période qui nous occupe ici.

de rendre compte des concepts de toutes les langues. S. Dodd proposait quant à lui une version simplifiée et standardisée de l'anglais, *le Model English*. Enfin, Dostert se déclarait en faveur du développement d'une syntaxe générale (*core syntax*). Toutefois aucune de ces idées, issues de conceptions trop diverses, peu étayées théoriquement⁴ ou bien ne trouvant pas d'écho dans les discussions en cours, n'a donné lieu à une véritable application, à l'exception du groupe britannique de Cambridge qui a développé une méthode sémantique par thesaurus dont Halliday a proposé une version multilingue en 1956. Seuls les Soviétiques ont véritablement réalisé des algorithmes fondés sur un modèle de langue intermédiaire.

Dans cet article, nous nous proposons d'examiner en quoi la question d'une langue intermédiaire, en tant que méthode de TA, rencontre, d'une part, un fort besoin de traduction du russe en diverses langues, lié au multilinguisme régnant dans l'Union, et d'autre part une tradition linguistique encore très vivante en URSS, celle de la linguistique comparée issue de Baudoin de Courtenay. La langue intermédiaire s'inscrit également dans la tradition de la standardisation de la langue russe promue par des espérantistes comme Drezen. Enfin, il convient de considérer une troisième source répondant à une demande conjointe de traduction multilingue et de productivité technique : c'est la machine de Trojanskij, premier projet d'automatisation de la traduction conçu en 1931, bien avant l'apparition des premières calculatrices électroniques en 1945. Autant de sources qui contribuent à la spécificité russe de la langue intermédiaire et qui vont conduire les premiers expérimentateurs à envisager la traduction à travers la mise en œuvre de caractéristiques communes à diverses langues. On verra également comment la réalisation de modèles de langue intermédiaire par Bel'skaja, Andreev ou Mel'uk est traversée par une réflexion théorique à la fois sur les algorithmes et sur le travail linguistique de recherche des caractéristiques communes aux langues nécessaires à la TA.

1. LE PROBLEME DU MULTILINGUISME

Le multilinguisme est, pour la Russie, un problème récurrent autant qu'ancien. La claire conscience de ce multilinguisme accompagnait déjà les expansions territoriales de Pierre le Grand, puis de Catherine II au XVIII^e siècle. Elle s'exprimait dans le recueil commandé par Catherine II à

⁴. Les préoccupations américaines étaient davantage orientées vers la création de grammaires universelles. Ainsi L. et A. Wundheiler exposaient la nécessité d'un langage universel logique pour la TA, inspiré des travaux de Carnap et d'Ajdukiewicz. Cependant, même si les premières recherches de Bar-Hillel sur les grammaires catégorielles à partir des travaux d'Ajdukiewicz et de Harris ont été effectuées dans le cadre de son groupe de TA au MIT, les grammaires formelles à visée universelle se sont développées dans un contexte de mathématisation de la grammaire dont toute préoccupation technologique était absente. Elles ne furent donc guère envisagées comme méthodes de TA.

Simon Pallas (1741-1811). Dans son *Avis au public concernant les vocabulaires comparés des langues de toute la terre* (1785)⁵, Pallas décrivait

L'Empire de Russie qui s'étend sur une grande partie de cette Asie, partie presque inconnue aux savants dans les temps antérieurs à Pierre le Grand, [et qui] contient sans doute plus de nations et de peuplades, de langues et de dialectes qu'aucun autre Royaume de la terre. L'espace très resserré du Caucase, habité par des peuplades peu nombreuses et très voisines entre elles, recèle plus de vingt-deux dialectes de huit ou neuf langues différentes. La Sibérie, plus vaste, en offre un plus grand nombre encore, et la seule presqu'île de Kamtchatka, dont la population, lors de sa découverte par les Russes, ne semblait que commencée, contenait neuf dialectes différents de trois langues hétérogènes.

Le problème de la définition d'une langue commune à cet empire se posait dans le même temps. Tout au long du XVIII^e siècle, s'est élaborée la codification de la norme de cette langue vulgaire commune, aux confins de la grammaire slavonne et des descriptions du russe vernaculaire qu'avaient produites les Étrangers. Il n'est pas le lieu ici de développer la perpétuelle tension entre la prescription d'un russe stabilisé et normé et la prise en compte d'un multilinguisme de fait, dont la vigueur est attisée par les événements actuels. Il importait pour notre propos d'évoquer une situation qui n'a fait que se perpétuer et que l'on peut résumer à ces quelques chiffres relativement récents :

Les résultats d'un recensement de la population mené peu avant l'éclatement de l'URSS (1979), révélaient que 28,1 % des habitants de l'Union Soviétique, soit 73 millions de personnes, étaient bilingues, pour 61 millions d'entre eux, soit 23,4 %, le russe était une deuxième langue⁶, — ce qui nous indique au passage que les sujets bilingues sont majoritairement des sujets non russes —. On comptabilisait 130 langues représentées, appartenant à 9 groupes différents : indo-européennes, caucasiennes, sino-tibétaines, mongoles, paléo-asiatiques, samoyèdes, tungouses-mandchoues, turkes, finno-ougriennes.

2. TRADITION LINGUISTIQUE COMPAREE ET PROTO-LANGUE

C'est Vjaeslav Ivanov, né en 1929, connu pour ses travaux sur l'indo-européen et le hittite, qui propose de réinvestir dans la réflexion sur la langue intermédiaire l'immense travail théorique et pratique mené par la linguistique historique et comparée. Ses propositions, contenues dans sa participation à la Conférence de Linguistique Mathématique de 1959 (*Théorie des relations entre les systèmes linguistiques et fondements de la linguistique historique et*

⁵. Cf. Traduction de Dariusz Adamski, dans Causat P., Adamski D., Crépon M. (1996), *La Langue, source de la nation ; messianismes séculiers en Europe centrale et orientale du XVIII^e siècle au XX^e siècle*, Liège, Mardaga.

⁶. Selon les résultats d'un recensement mené en 1979, cités par G.A. Zograf (1990), in *Lingvistijskij enciklopedijskij slovar'*, Moskva, Sovetskaja Enciklopedija.

comparée), connaîtront un vaste écho ainsi qu'une influence certaine, sur les conceptions d'Igor Mel'uk notamment, dont Ivanov fut le directeur de thèse.

L'idée en est que les méthodes de comparaison des faits de langue sont largement éprouvées, mais qu'une métathéorie capable de décrire la structure formelle de la comparaison, une *théorie des relations des systèmes linguistiques* reste à construire. Cette théorie serait d'une importance capitale pour nombre de disciplines scientifiques nouvelles, de l'anthropologie structurale à la traduction automatique, en passant par la théorie de l'information.

Elle s'appuierait sur une langue abstraite, nommée proto-langue, qui ne serait pas représentative d'un état antérieur d'une langue donnée, mais qui en constituerait le système abstrait, au sens où celle-ci recèlerait toutes les potentialités communes à un groupe de langues, que ces potentialités soient exprimées ou non. La proto-langue serait obtenue grâce à la méthode comparatiste, par reconstruction interne d'un sous-système, compatible avec le système dans son ensemble. C'est cette proto-langue que propose Ivanov comme langue intermédiaire pour la TA.

Il s'inscrit ici dans l'héritage direct de Baudoin de Courtenay [1845-1929], puisqu'il fait référence explicitement à la théorie de la mixité de toutes les langues que développait Baudoin⁷. Ivanov retient un système linguistique différencié, c'est-à-dire composé d'une série de systèmes dont chacun peut être rapporté à un autre système : "La thèse de Baudoin de Courtenay sur le caractère mixte de toutes les langues se réduit à l'affirmation d'une possibilité de différenciation de tout système de langue."

Ivanov reprend en outre deux autres idées fortes de Baudoin de Courtenay, à savoir le caractère artificiel de la langue mère recréée par le linguiste et sa conception de la "langue nationale" comme système abstrait. Prenant ses distances avec les explications génétiques et organiques, Baudoin posait que ces langues, restituées par la science, ne sauraient être prises pour des langues

⁷. L'article intitulé "O smeannom xaraktere vsejazykov" [Du caractère mixte de toutes les langues] constitue la leçon inaugurale du Cours de *grammaire comparée des langues slaves en relation avec les autres langues aryoeuropéennes* prononcé le 21/09/1900 à l'Université de Saint-Petersbourg, publié dans *urnal Ministerstva Narodnago Prosveenija, Tom CCCXXXVII, 1901* [Revue du ministère de l'Instruction publique]. Baudoin y développe l'idée que la mixité des langues, mixité géographique et historique, est un phénomène incontournable et concernant toutes les langues. Mixité géographique, car les situations de confins créent les conditions de la polyglossie, mixité chronologique, car la langue ancienne peut s'incruster dans la langue moderne, ou même exercer une influence sur la langue moderne (dans le cas par exemple de la langue religieuse).

La mixité des langues joue dans deux sens : elle apporte dans une langue donnée des éléments provenant d'une autre langue (lexique, constructions syntaxiques, morphologie, prononciation...), elle est responsable de l'aplanissement de la "force de différenciation", qui existait entre divers éléments d'une langue particulière. Au titre de ses effets, on compte l'homogénéisation, l'aplanissement, la perte du caractère flexionnel, l'extension des formes analytiques...

réelles, mais pour des “ faits scientifiques, obtenus par déduction ”⁸. Il préconisait des analyses typologiques pour faire apparaître les caractéristiques communes des langues affiliées.

Pour Baudoin de Courtenay, les langues recréées sont des observatoires, ce qui entre en cohérence également avec sa vision de la langue nationale, vue comme une abstraction des faits typiques, conformes à la loi générale et essentiels aux langues individuelles. “ Elle se présente comme une pure abstraction, un édifice d’unification. ” Les langues artificielles, nécessitant un travail conscient de type métalinguistique (langue littéraire, langages des métiers...) apparaissent comme autant de sous-systèmes de la langue nationale.

La proto-langue que propose Ivanov comme langue intermédiaire est bien dotée des particularités décrites précédemment. Il s’agit bien, comme c’était le cas chez Baudoin de Courtenay, non d’un modèle a priori, mais d’un système second, d’un modèle intégrant les caractéristiques potentielles des différentes langues étudiées au préalable, d’un modèle évolutif capable de s’ouvrir à de nouvelles langues, à condition de suivre un protocole d’intégration.

Nous reviendrons plus loin sur quelques mises en œuvre de cette proto-langue dans des modèles de langue intermédiaire, et nous retiendrons pour l’instant qu’Ivanov voit ici un moyen de revivifier la réflexion issue du comparatisme, et le vaste matériau linguistique produit par ce courant, très puissant en Russie, mais aussi dans cette fin des années 50, très déconsidéré d’avoir servi de nourriture intellectuelle au marrisme et à sa théorie de la stadialité⁹.

3. DREZEN ET L’INTERNATIONALISATION PROPORTIONNELLE DES LANGUES, 1937

Un autre courant extrêmement fort avant-guerre, — auquel Baudoin de Courtenay¹⁰ avait apporté également une participation éminente, puisqu’il contribua à la popularisation de l’espéranto —, c’est l’élaboration des langues artificielles. La situation révolutionnaire, l’idée d’une exportation de la révolution à la planète entière, ont promu les travaux sur les langues inter-

⁸. Cf. l’article “ Nekotoryja obija zameanija o jazykovedenii i jazyke ” [*Quelques remarques générales sur la linguistique et la langue*], leçon inaugurale prononcée pour l’ouverture de la chaire de grammaire comparée des langues indo-européennes (17/29 décembre 1871, Université de Saint-Petersbourg). Des extraits en ont été traduits par Dariusz Adamski dans “ Baudoin de Courtenay et la linguistique générale ”, *LINX* 23, *Traduction de textes peu ou mal connus*.

⁹. Ivanov insiste sur le fait que “ les méthodes de formation des modèles abstraits communs aux langues apparentées doivent être mises en évidence par des recherches typologiques dans le domaine de l’anthropologie structurale et ne pas être postulées a priori ”.

¹⁰. On rappellera que le manuel fondateur de l’espéranto rédigé par Lejzer Ludwik Zamenhof, parut pour la première fois en russe en 1887.

nationales. C'est le cas de la langue cosmique, langue de communication internationale, dite *langue AO*, qui fut créée par Gordin en 1920, dans le sillon du phénomène de *dimension cosmique* qu'est la révolution (cf. Kuznecov 1995). Dans les années 30, la construction des langues auxiliaires est encore à l'ordre du jour, ravivée par la forte demande terminologique qui accompagne le développement des techniques. Ces travaux retentiront également sur les projets de langue intermédiaire.

Le courant liant la standardisation de la langue scientifique et technique et les langues internationales artificielles est représenté, tant au niveau scientifique qu'institutionnel par Ernest Karlovi Drezen, bolchevik de premier plan, responsable de la standardisation de la langue auprès du Commissariat du peuple à la guerre et espérantiste éminent : Membre, puis Président, de la société *Espéro*, membre de la société *Kosmoglott*, avec Baudoin de Courtenay, entre 1915 et 1918, il en devient secrétaire perpétuel en 1925, il est également secrétaire général de l'Union des espérantistes, membre du *Jazykfront* (Front de la Langue).

Auteur prolifique, dont l'œuvre sera brutalement interrompue en 1936 par son arrestation, puis son exécution (cf. Tatarinov 1993), Drezen affronte deux problèmes principaux : la nécessité d'arriver à une langue technique standardisée et le rapport de cette langue aux langues nationales d'une part, à la langue internationale d'autre part.

Dans son dernier ouvrage, paru en 1936, intitulé *L'Internationalisation de la terminologie scientifique et technique, Histoire, actualité et perspectives*, Drezen étudie de près les différentes langues universelles ayant existé ou existant. Il critique l'usage du latin comme langue scientifique contemporaine (Drezen, *op. cit.*, 12), et réfute de la même façon toute langue artificielle ou semi-artificielle obtenue à partir d'une langue nationale, ce pour des raisons linguistiques (elle sera accessible surtout aux locuteurs de cette langue nationale, ou à la rigueur des langues proches), mais aussi idéologiques¹¹. Dans la société capitaliste, les contradictions inhérentes à la société concernent également les langues : les unes rechercheront l'hégémonie, les autres à survivre et se maintenir.

Drezen développe des arguments en faveur d'une langue internationale créée de toutes pièces. En effet, du point de vue strictement terminologique, l'imprécision, la polysémie des notions et des termes scientifiques et techniques repris à des langues naturelles sont pour lui rédhitoires. S'appuyant sur l'inadéquation de termes tels *Force-Cheval*, *loadinaja sila*, *Pferdestärke*, *Horse-Power*¹², il souligne la complication qu'il y a à établir des dictionnaires

¹¹1. Drezen cite l'exemple du latin sans flexion, l'Interlingua de Peano.

¹²2. Tout d'abord, la signification technique du terme (soit 75 kilogrammètres/sec = 0,736 KW) est en décalage avec le mode de désignation métaphorique, mais en plus, le *Horse-Power* des Anglais excède cette force métrique de 1 %. (Drezen, *op. cit.*, 24).

techniques bilingues, les énormes difficultés d'harmonisation dès que l'on passe à une troisième langue, sans même envisager le cas des dictionnaires polyglottes.

Selon Drezen, seule une langue internationale artificielle peut être univoque. La tâche reste à accomplir. Or, si l'on tient compte du fait que la science et la technique se développent à une vitesse accélérée, et que les besoins en terminologie sont parallèlement accrus, il y a urgence à proposer des voies d'unification. C'est le but que se fixe Drezen, qui œuvre au sein des instances internationales, telles que l'*Association Internationale de Normalisation*, dans laquelle il tente de faire prévaloir les idées soviétiques, nourries des idées internationalistes qui avaient déjà guidé Mendeleev, et que Drezen cite à ce propos :

Toute personne qui comprend que viendra le jour d'un rapprochement étroit des peuples se doit de préparer le lien, aussi solide que possible, qui les unira.

L'homogénéisation des classifications scientifiques et techniques et de la terminologie afférente constitue un jalon vers ce rapprochement ultime.

Si, pour Drezen, l'idée forte d'une normalisation linguistique serait de supprimer du langage technique toute équivocité dans les termes, la synonymie, l'homonymie, les ambiguïtés, la *pseudo-internationalisation* de certains termes, il reste réaliste et prône une *internationalisation proportionnelle* de la langue technique, une mise en place progressive, par le biais notamment des termes nouveaux, de la future langue mondiale, dans son esprit, l'espéranto. À terme, l'évolution de la société menant à une société sans classe rejoindra l'évolution de la langue qui se sera progressivement internationalisée. Il s'agit d'un horizon lointain, vers lequel doivent tendre les décisions fragmentaires prises dans le domaine de la standardisation linguistique.

Les propositions qu'il fait en ce sens devant l'*Association Internationale de Normalisation de Stockholm* en 1934 recevront un certain écho, puisque la Conférence confie à l'URSS le soin d'avancer des propositions nouvelles sur ces questions (cf. résumé français et espéranto, Drezen 1936, 97).

Le débat sur les langues internationales est encore vif, y compris au niveau institutionnel. Les problèmes concernés sont tout autant la communication scientifique, les rapports entre les scientifiques eux-mêmes, les immenses besoins terminologiques qui accompagnent le développement des sciences et techniques. Par ailleurs, on assiste à un important effort de traduction en russe de textes techniques et scientifiques étrangers. Il existe donc à cette époque en URSS une demande accrue, liée aux besoins de terminologie dans les différentes langues de l'Union, ainsi qu'une nécessité d'opter pour une langue de communication scientifique commune.

Nous remarquerons que le terme *force-cheval* employé par Drezen ne semble pas attesté en français. C'est le terme *cheval-vapeur* qui est attesté depuis 1922.

Ce sont ces mêmes problèmes que tente de résoudre l'ingénieur P.P. Smirnov-Trojanskij (1894-1950) avec sa machine à traduire. Partant du constat "qu'il est impossible de traduire par d'anciennes méthodes la masse sans cesse accrue des textes", il propose d'élargir à la traduction la mécanisation qui a déjà fait ses preuves dans maints domaines d'application.

4. LA MACHINE DE TROJANSKIJ, 1931

Les objectifs de Trojanskij sont donc pris dans l'air du temps, et c'est avec certitude qu'il fustige le réflexe passéiste qui consiste à nier l'importance de la mécanisation et de la standardisation dans le nouveau mode de vie ; celles-ci sont pour lui inéluctables, preuve en est les nombreux domaines dans lesquels elles sont appliquées. L'analogie avec la mécanisation des campagnes n'est pas anodine : "On peut fort bien semer le grain à la main. Mais il existe des semeuses mécaniques tirées par des tracteurs". La "Machine pour la production automatique des traductions imprimées en plusieurs langues, ne nécessitant qu'une mise au point littéraire", à laquelle il travaille dès 1931, sera décrite dans un mémoire daté du 29 juillet-3 octobre 1933¹³.

Comme base d'automatisation de la traduction, Trojanskij postule une structure commune à toutes les langues, au delà de leur diversité ; entre la langue-source et la langue-cible, il suggère de passer par un texte intermédiaire, qu'il dit *logique*, qui consiste en fait en une suite de mots apparaissant à leur forme grammaticale de base, comme autant d'entrées de dictionnaire qui seront ensuite traduites : "J'ai pris pour point de départ l'existence d'une structure logique unique dans toutes les langues : un sujet reste toujours un sujet, un verbe — un verbe et un complément — un complément etc. Cette unité de construction logique crée un rapprochement entre les langues, malgré la diversité de leur structure, de leur grammaire et de leur vocabulaire. Elle rend les difficultés surmontables."

Dans le système de Smirnov, un traducteur obtient, à partir d'un texte en langue source A, un texte en langage *logique* A 1, que la machine traduira automatiquement en langage *logique* B 1, grâce à un dictionnaire qui traduit automatiquement les entrées. Le deuxième texte en langage logique ainsi obtenu sera ensuite arrangé manuellement par un rédacteur en langue cible B, selon le modèle ci-dessous d'une traduction de l'allemand simultanément en russe et en français¹⁴.

¹³. Cf. (Hutchins 1993) pour la description du fonctionnement de cette machine, et pour sa représentation établie sur la base d'un schéma figurant dans le brevet de Trojanskij, datant de 1933.

¹⁴. A est la langue source, B et C les langues cible ; A1 (résultat de l'analyse), B1 et C1 (entrée de la synthèse) sont les textes en langage dit logique, ils sont suivis d'une marque de reconnaissance du verbe (as) et du nom (o). Selon Hutchins (1993), qui a eu accès à l'ensemble des travaux publiés de Trojanskij, ces formes logiques sont empruntées à l'espéranto.

TRADUCTION AUTOMATIQUE EN URSS107

| A | A1 | (B1 + C1) | | (B + C) | |
|-------------|-------------------|----------------|-----------------|-----------|------------|
| Das Bild | Das Bild - o | rfnbyf - o | Le tableau -o | rfnbyf | Le tableau |
| der Welt | die Welt de - o | vbh de -o | le monde de -o | vbhfx | du monde |
| zeigt, | zeigen - as | gjrpsdfnm - as | montrer - as | gjrpsdfn, | montre |
| wie | wie | rfr | comment | rfr | comment |
| die Materie | die Materie - o | vfthbz - o | la matière - o | vfthbz | la matière |
| sich bewegt | sich bewegen - as | ldbufnmcz - as | se mouvoir - as | ldb;ncz, | se meut |
| wie | wie | rfr | comment | rfr | comment |
| die Materie | die Materie - o | vfthbz - o | la matière - o | vfthbz | la matière |
| denkt | denken - as | vsckbnm - as | penser - as | vsckbn | pense |

Pour Trojanskij, le grand intérêt de la méthode réside dans le fait qu'il s'agit d'une méthode de traduction *monolingue*, chaque traducteur pouvant ne connaître qu'une langue et une seule : il met ainsi le travail de traduction à la portée de tous les citoyens, puisque ce travail s'apparente aux exercices des écoliers du cours élémentaire qui décomposent un texte de leur langue maternelle "en parties de propositions et en parties de discours". La machine pouvant traduire le langage logique source en plusieurs langages logiques cible, la restitution en langue "nationale" à partir de la langue intermédiaire peut être effectuée par le destinataire du message.

Selon Trojanskij, les avantages économiques de la méthode sont nombreux : outre la possibilité de faire appel à des traducteurs non qualifiés, elle permet de publier des textes en langage logique directement, pour une traduction rapide dans plusieurs langues cible, de faciliter les échanges de correspondances (notamment scientifiques) entre personnes parlant des langues différentes, ainsi que les traductions de livres et revues, de traduire simultanément en plusieurs langues. Plus le nombre de ces langues est élevé, plus le prix de revient est bas. Selon Trojanskij, qui se garde de spécifier sa méthode de calcul, on peut réaliser des économies allant jusqu'à 99 %. C'est l'embryon de langue intermédiaire qui permet de comprimer les coûts. Il réfute d'ailleurs le fait que sa machine soit uniquement un dictionnaire. Pour lui, il y a réellement langage intermédiaire en l'espèce de ces marqueurs de verbe et de nom¹⁵.

Trojanskij croit en l'avenir de sa machine à traduire ; il travaille aux applications jusqu'en 1939, puis après un arrêt forcé, il revient à son projet en 1947. Ce sont les pionniers de la TA en URSS, Bel'skaja, Panov et Korolev qui publieront en 1959 une édition commentée de l'ouvrage que Trojanskij avait

Trojanskij prévoyait à terme d'automatiser le travail du rédacteur qui opère à partir de la langue source, en construisant une machine capable d'effectuer le travail pour le rédacteur nommé "le logicien". Cette machine devait pouvoir reconnaître des traits morphologiques ou syntaxiques permettant de distinguer un verbe d'un substantif.

¹⁵. En ce sens, la machine de Trojanskij se distingue de la méthode de traduction interlinéaire (cf. Auroux 1979, 194-196).

rédigé en 1947 : *De la machine à traduire, construite sur la base d'une méthode monolingue*. Ils insistent sur la précocité de ce travail, ainsi que sur l'ampleur des problèmes que l'auteur avait abordés, tentant d'instaurer ainsi les Russes comme pionniers dans le domaine de la TA.

5. LES PREMIERES EXPERIENCES DE TA : PRESENTATION DES CENTRES DE RECHERCHES (1954-1960)

Ces recherches se développèrent dans cinq principaux centres, quatre à Moscou et un à Leningrad (*cf.* Harper 1960, Hutchins 1986).

À Moscou, deux centres au sein de l'Académie des Sciences font de la TA. D. Ju. Panov, qui avait assisté à la première démonstration de TA sur calculatrice électronique à New-York en 1954, dirigeait le groupe de TA au sein de l'Institut de mécanique de précision et de calcul mécanique dirigé par l'académicien D.S. Lebedev. Né en 1904, docteur ès-sciences techniques, Panov est spécialiste des problèmes mathématiques posés par la technique moderne. Son groupe, préconisant l'utilisation du russe comme langue intermédiaire, comprenait une philologue, I.K. Bel'skaja, spécialiste d'anglais, et des ingénieurs électroniciens tels que L.N. Korolev, A.I. Muxin et S.N. Razumovskij.

Le centre de TA de l'Institut de mathématiques Steklov de l'Académie des Sciences de Moscou (MIAN) était dirigé par A.A. Ljapunov (1911-1973), docteur ès-sciences. Spécialiste de balistique des fusées, il avait un poids politique et militaire important qu'il utilisa pour promouvoir les recherches en TA. Ce groupe, à l'orientation fortement théorique, comprenant également la mathématicienne O.S. Kulagina et la philologue T.N. Molonaja (nées toutes deux en 1932), contribua de façon spécifique au développement de la théorie de la programmation. À l'Institut de Linguistique de l'Académie des Sciences de Moscou, le linguiste I.A. Mel'uk, né en 1932, développa, à partir d'un algorithme de traduction russe-hongrois, l'idée d'une langue intermédiaire. Il avait développé auparavant en collaboration avec Kulagina un système de TA français-russe.

Outre ces grands instituts de l'Académie des Sciences, un séminaire hebdomadaire animé par V. Ju. Rozencvejk au sein de l'Institut Pédagogique d'État des Langues étrangères de Moscou, l'Institut Maurice Thorez, regroupa les pionniers autour des premières discussions sur la TA. Cet institut accueillit des recherches focalisées plus particulièrement sur la théorie sémantique, menées en particulier par I.I. Revzin (1923-1974). En 1958 le bulletin de l'Institut devint la revue de TA de l'URSS : *Mainnyj Perevod i Prikladnaja Lingvistika*.

Alors qu'à Moscou, la TA se développe à l'Académie des Sciences¹⁶, à Leningrad, c'est au sein de l'université, au Laboratoire expérimental de traduction par machine dirigé par N.D. Andreev et rattaché à l'*Institut de Mathématiques et de Mécanique*, que se développent ces recherches. Spécialisé dans l'examen d'un nombre considérable de langues différentes, ce groupe développe les bases d'une langue intermédiaire, fondée sur les traits communs aux différentes langues¹⁷.

Ces grands centres sont les plus connus, mais de nombreux autres groupes sont aussi impliqués dans la recherche en TA. On en veut pour preuve le premier congrès sur la TA organisé en 1958 à Moscou qui rassembla quelques 340 participants issus de 79 institutions différentes. Lors de ce Congrès, la langue intermédiaire constituait un thème important et largement discuté dans les débats.

6. LES PROBLEMES THEORIQUES

Lorsqu'ils entreprennent les premières expériences de TA, et en accord avec la tradition linguistique russe dont ils sont issus, les chercheurs vont mettre au premier plan l'idée que la traduction d'une langue à une autre implique la mise en œuvre d'invariants linguistiques et conceptuels communs à toutes les langues. De plus, les linguistes voient dans la TA, au-delà d'objectifs purement pratiques, un lieu d'investigation totalement nouveau pour l'étude des langues, parmi lesquels l'examen des relations entre les expressions de la même idée dans différentes langues (*cf.* Kuznecov, Ljapunov, Reformackij 1956).

Par ailleurs, la réflexion théorique sur la traduction et la traductibilité des langues, quasiment absente des travaux anglo-américains et européens en général — l'ouvrage d'Eugene A. Nida *Toward a science of Translating* n'a été publié qu'en 1964 —, est en revanche intense et continue en Russie depuis le XVIII^e siècle¹⁸. Dans les années 50, un important débat anime les théoriciens de la traduction à Moscou à partir de la parution en 1953 de l'ouvrage de A.I. Fëdorov (né en 1921) "Introduction à la théorie de la traduction", qui entreprend de poser les bases objectives d'une science de la

¹⁶. Une des raisons de cette localisation tient au fait que Mel'uk et Ivanov avaient été exclus de l'Université de Moscou en 1957 pour avoir pris position en faveur de Pasternak.

¹⁷. G.S. Cejtin (né en 1936) et S. Ja. Fitjalov ont été rattachés un temps à ce groupe, avant de créer, leur propre groupe, l'*Intelligent Systems Laboratory*, encore existant, au sein du même Institut (communication personnelle de G. Cejtin).

¹⁸. On pense par exemple à la célèbre colère de Pierre le Grand lorsqu'il eut en main la traduction en russe du *Traité des fortifications* (Istinnyi sposob ukreplenija gorodov, izdannyi ot slavnogo inenera Vobana [...] Sankt Peterburgskaja Tipografija, 1724), dont la traduction littérale rendait impossible l'édification de fortifications. Il imposera alors que les traductions visent principalement la transmission d'informations directement utilisables.

traduction. Il est intéressant de noter que l'ouvrage de Fëdorov n'était pas inconnu des expérimentateurs en TA (*cf.* Bel'skaja 1958).

Les Russes adoptent une position critique à l'égard de la conception américaine des langues et de la traduction, représentée par le Memorandum de Weaver publié en 1949. Bien qu'ayant suivi, au début du moins, une démarche empirique de traduction mot-à-mot, en tout point semblable à celle de l'Université de Georgetown, Panov (1956) et son groupe critiquent explicitement la position de Weaver selon laquelle la langue est un code et la traduction un décodage. En codant et en décodant, dit Panov, on ne change que la forme externe des mots sans toucher à la langue elle-même. Une telle tâche peut être résolue par des méthodes formelles. Mais en traduisant, on a affaire à des langues, c'est-à-dire, et c'est la position défendue par Ivanov, à des systèmes complexes et subtils permettant d'exprimer des pensées, ancrés dans l'histoire et les pratiques sociales. La traduction implique de transférer du sens d'une langue à l'autre et ne peut se traiter comme un changement de code. L'analyse des structures linguistiques ne peut donc à elle seule résoudre les problèmes de traduction. Il faut aussi un système sémantique formel.

Des considérations linguistiques sont aussi à l'origine de la langue intermédiaire. La traduction de l'anglais vers le russe ne peut utiliser aussi largement l'analyse des flexions¹⁹ qu'une traduction à partir du russe. Il faut, par exemple, envisager des unités de traduction différentes du mot graphique, unité de la traduction mot-à-mot — *cf.* Molonaja 1957 qui propose une analyse en constituants immédiats issus de Fries —. Les propriétés spécifiques des langues vont aussi imposer d'autres contraintes de traduction. Alors que, dans la traduction de langues comme le russe, l'anglais, le français ou l'allemand, il était possible de considérer la différence d'ordre des mots comme négligeable, ce n'est pas possible dans la traduction hongrois-russe. C'est une des raisons qui inciteront Mel'uk (1958) à envisager l'utilisation d'une langue intermédiaire.

De plus, l'idée de langue intermédiaire va se trouver impliquée dans deux types de préoccupations technologiques. La première est d'ordre algorithmique : c'est l'indépendance des algorithmes d'analyse et de synthèse. La seconde est liée à une conception de la TA, non comme une fin en soi, mais comme la première étape d'un programme plus général visant à utiliser les machines électroniques comme des outils auxiliaires de la pensée humaine. Deux autres points, qui seront au cœur des premières tentatives de création d'une langue intermédiaire, concernent la comparaison des langues : (i) il s'agit de savoir s'il est préférable de définir des relations formelles entre une langue artificielle et chacune des langues naturelles considérées ou entre deux

¹⁹. Dans un article de 1957, Molonaja, Purto, Revzin et Rozencvejk tiennent le russe pour une langue riche en possibilités formelles, grâce notamment à ses terminaisons casuelles qui le rendent intéressant comme langue source pour la TA.

langues naturelles ; (ii) dès lors que la langue intermédiaire repose sur les invariants des langues considérées, que faire de la variation ?

Langue intermédiaire et théorie des algorithmes

Dès 1954, Léon Dostert (Locke et Booth, 1955), directeur du groupe de Georgetown et instigateur de la première démonstration sur ordinateur, expose l'intérêt d'une langue intermédiaire (une syntaxe noyau) pour la TA multilingue. L'argument est essentiellement économique : dans le cas de la traduction binaire on a besoin de $N(N-1)$ algorithmes pour N langues à traduire ; grâce à une langue intermédiaire, il n'en faut plus que $2*N$.

Mais au-delà de cette raison économique, reprise d'ailleurs par les Soviétiques, l'utilisation d'une langue intermédiaire a des causes plus profondes. Pour Mel'uk (1959), l'utilisation d'une langue intermédiaire sert non seulement à économiser le nombre des algorithmes mais à les standardiser. Il insiste sur la nécessité d'élaborer une théorie générale de la compilation des algorithmes de traduction, à laquelle doivent participer les mathématiciens. Celle-ci a pour objectif de mettre au point une forme universelle d'algorithme de traduction et de généraliser les règles d'applications aux langues concrètes²⁰.

L'indépendance des algorithmes d'analyse et de synthèse

L'attention des Soviétiques portée au processus de traduction lui-même, c'est-à-dire au transfert de sens entre langue source et langue cible, les conduit à distinguer, sur le plan algorithmique, le processus d'analyse de celui de synthèse²¹. Les russes (cf. Panov, Ljapunov et Muxin 1956) reprochent aux travaux américains, orientés vers la traduction mot à mot, la liaison trop étroite entre dictionnaire et procédure de traduction, ce qui favorise la seule analyse de la langue source au détriment de la synthèse en langue cible. Ils critiquent tout

²⁰. Ces préoccupations font écho à un certain flou sur l'utilisation des termes *algorithme* et *langue intermédiaire* par les expérimentateurs de TA (cf. Andreev, Ivanov, Mel'uk 1960 et Mel'uk 1961). Selon certains, la langue intermédiaire est directement un algorithme. Pour d'autres, elle est indépendante. Ainsi Yngve soutient que la plupart des méthodes utilisent de fait une langue intermédiaire dans leurs algorithmes : dans le dictionnaire français-anglais du groupe de Birbeck College de Londres, l'équivalent anglais du mot français correspondant n'est pas directement stocké, c'est son adresse qui est stockée, et c'est cette adresse qui constitue dans ce cas, selon Yngve, une langue intermédiaire.

²¹. Yngve suggéra très tôt l'utilisation d'une langue intermédiaire et la dissociation analyse-synthèse et cette proposition était connue des soviétiques (cf. Mel'uk 1961). Toutefois aucun algorithme concret ne fut réalisé et il préféra développer une autre méthode, la méthode dite de *transfert* fondée sur une analyse syntaxique des langues source et cible. Lors du premier congrès soviétique sur la TA qui s'est tenu à Moscou en mai 1958 (cf. Nikolaeva 1958), I.I. Revzin réinsère dans la tradition russe l'idée d'une approche séparée entre analyse et synthèse en la rapprochant de la distinction entre grammaire active (la synthèse) et grammaire passive (l'analyse) effectuée par l'académicien erba à propos de l'enseignement des langues.

spécialement le fait que les instructions chargées d'effectuer l'analyse grammaticale soient directement associées aux mots contenus dans le dictionnaire et préconisent au contraire un dictionnaire indépendant du programme de traduction. Ce traitement "déclaratif" du dictionnaire, opposé au traitement "procédural" des Américains a, selon eux, deux avantages : le dictionnaire est rendu aisément modifiable et la synthèse en langue cible est indépendante de l'analyse de la langue source.

Langage d'information et langue intermédiaire

Pour la plupart des chercheurs soviétiques (*cf.* Mel'uk 1959), la TA n'est qu'une première étape d'un programme plus général de traitement de l'information, de son extraction et de son stockage à partir de textes écrits dans différentes langues, tel qu'éditer, faire des résumés, fournir des références bibliographiques etc., et pour lequel il faut un langage spécial. Le langage d'information ainsi conçu a un usage documentaire. La TA fait donc partie d'un projet plus vaste mobilisant toutes les énergies²².

En ce qui concerne la TA, les approches seront différentes selon les conceptions linguistiques et algorithmiques présidant au transfert d'information : entre deux langues naturelles ou entre une langue naturelle et une langue artificielle.

Panov sera ainsi amené à distinguer nettement la langue intermédiaire, qui doit être une langue naturelle, d'une langue d'information artificielle²³. Pour Mel'uk et Andreev, au contraire, le langage d'information peut aussi servir de langue intermédiaire pour la TA.

7. LE RUSSE COMME LANGUE INTERMEDIAIRE

Dans les toutes premières expériences menées par le groupe de Panov, la langue intermédiaire, consacrée exclusivement à la confection d'algorithmes multilingues, doit être une langue naturelle. Panov (1956) écarte les langues artificielles, telles que les langues internationales a posteriori comme l'espéranto, l'interlingua, le volapük, qu'il tient pour incapables de transmettre la richesse des langues naturelles. Il écarte également l'idée de construire une langue spécifique adaptée à la traduction et propose le russe comme langue intermédiaire.

²². Une résolution du 20^e Congrès du PCUS de 1956 a promu de façon explicite la TA comme projet scientifique prioritaire.

²³. Pour Panov, la création d'une langue d'information est directement liée à la TA et à la résolution des ambiguïtés sémantiques. Cette langue, qui doit être univoque, rappelle la langue univoque de Drezen. Couplée avec l'utilisation de micro-glossaires spécialisés, elle est destinée à décrire le contenu d'un article scientifique dans ses grandes lignes, de façon résumée et sans ambiguïtés. Toutefois le groupe de Panov n'a jamais réalisé de langue d'information qui est restée à l'idée de projet.

| | | | | | | | |
|-------|--|--|---------------|-------------------------|-----------------|-----------------|---|
| russe | | | <i>teper'</i> | <i>rassmo- trim</i> | <i>razlinye</i> | <i>operacii</i> | , |
|-------|--|--|---------------|-------------------------|-----------------|-----------------|---|

| | | | | | | | |
|--|-----------------------------------|--|--|--|---------------------|-------------------------|---|
| anglais | <i>which</i> | <i>may</i> | <i>be</i> | <i>performed</i> | <i>upon</i> | <i>vectors</i> | . |
| infos grams sur l'anglais | adj conj | verbe prédictat modal | verbe | verbe | prép | nom | |
| infos grams sur le russe | nom conj suj acc plur | forme impers. sauter l'acc vers (* - -1) | prédictat inf sauter l'acc vers (* - -1) | prédictat, inf inverser le sujet en obj dir. prend l'acc | prend l'instrum. | masc. plur. instrum. | |
| russe | <i>kotorye</i> | <i>mono</i> | | <i>vypolnit'</i> | <i>nad</i> | <i>vektorami</i> | . |

Les trois premières lignes représentent le résultat de la phase d'analyse, la quatrième, le résultat de la phase de synthèse. À partir de l'analyse de la phrase anglaise, on obtient deux types d'informations grammaticales : les caractéristiques invariantes de la forme anglaise (partie du discours, genre, nombre, personne) mais surtout des informations grammaticales sur la forme russe correspondante. Celles-ci peuvent être des informations sur la morphologie mais ce sont aussi des instructions pour la construction de la phrase ou des syntagmes : par exemple la rection d'un verbe (+ accusatif) ou d'une préposition (+ instrumental), ou encore la façon de traduire un passif (prévoir l'inversion du sujet et de l'objet). Ce peut être aussi des instructions de déplacement sur la séquence, de strictes instructions de passage si l'on peut dire.

Cette conception de l'analyse devrait assurer, selon les auteurs, l'indépendance des algorithmes d'analyse et de synthèse. Une fois l'analyse faite, il n'y a plus besoin de revenir au texte source. La synthèse en russe, à partir de la traduction mot à mot et des informations grammaticales sur les formes russes, devient totalement indépendante de la phase d'analyse et du texte source. Les règles de synthèse ne dépendent pas de la langue source mais sont fondées sur du matériau linguistique russe, sur les règles de formation des mots russes et de réarrangement de l'ordre des mots.

Ces informations sur les formes russes, déduites de l'analyse des formes anglaises, qui sont en fait des instructions pour la traduction attachées au lexique, constituent donc pour les auteurs la langue intermédiaire.

Bien que les auteurs ne s'expliquent pas sur la façon dont cette déduction s'effectue automatiquement — ce qui somme toute est la clé du processus —, et qu'un tel positionnement du russe comme langue intermédiaire réduit les variations entre langues à quelques traits de morpho-syntaxe qui peuvent être mis en commun entre le russe et une autre langue à des fins de traduction, on peut toutefois avancer l'hypothèse que cette attention portée à ce qui est identique d'une langue à l'autre, et à ce qui les différencie — une langue peut

présenter une forme absente dans l'autre — constitue un premier pas permettant d'échapper aux impasses du mot à mot.

8. CONSTRUCTION D'UNE LANGUE INTERMEDIAIRE ARTIFICIELLE

La construction d'une langue intermédiaire artificielle pose deux types de questions : (i) le rapport d'une telle langue avec les langues artificielles auxquelles, on l'a vu, la tradition linguistique russe n'est pas étrangère ; (ii) la question de la représentation sémantique des invariants linguistiques, et de leur traitement automatique dans un processus de traduction.

Deux tentatives principales de création de langues intermédiaires artificielles ont été proposées par les Soviétiques. Celle d'Andreev, dont la méthode de construction s'apparente à celles des langues internationales, est issue de l'idée de langue univoque d'information promue par Drezen. Celle de Mel'uk est un langage de représentation sémantique abstrait, issu de la proto-langue d'Ivanov, qui s'apparente davantage aux langues universelles a priori, et qui est à l'origine de son modèle sens-texte.

Langue intermédiaire et congruence des langues

Pour Andreev, la construction d'une langue intermédiaire repose sur un traitement statistique des invariants linguistiques et sur une conception des langues comme faits sociaux. C'est une langue artificielle avec son lexique, sa morphologie et sa syntaxe propres, dont seules font partie les structures suffisamment distribuées dans les langues naturelles examinées, pondérées, par le nombre de locuteur et le nombre de textes déjà traduits. On reconnaît là la position statisticienne se prévalant de neutralité qui, au XIX^e siècle, présidait au choix d'une langue internationale a posteriori (Rasmussen 1995)²⁵, ainsi qu'une tendance bien connue en URSS à la planification chiffrée.

À travers le terme même de *hiéroglyphes*, par lequel Andreev nomme les éléments de sa langue intermédiaire, on reconnaît un des vieux mythes de construction d'une langue parfaite des images, de même que certains projets de pasigraphie du XIX^e siècle (cf. Salmon 1992).

Andreev propose de calculer pour chaque langue un indice de congruence à partir du nombre de hiéroglyphes concordants (le hiéroglyphe de la langue d'entrée est le même que celui de la langue de sortie) et non concordants qu'elle comporte, et de la pondération qui lui est affectée.

Soit la phrase anglaise suivante et sa traduction en hiéroglyphes :

| | | | | | |
|------------|------------|------------|----------|---------------|-------------|
| <i>the</i> | <i>sun</i> | <i>is</i> | <i>a</i> | <i>yellow</i> | <i>star</i> |
| φ5τ4 | + δ1 | + φ1 φ8 τ5 | + φ6 τ4 | + δ2τ2 | + δ3 τ1 |

²⁵. Selon une conception qui tenait les langues pour des faits sociaux, le choix d'une langue commune résultait d'une analyse scientifique des langues en Europe en fonction des masses de population qui la parlaient et des zones géographiques où elle était à l'œuvre.

Les hiéroglyphes sont des symboles écrits sous forme de suites de lettres et de chiffres.

- les hiéroglyphes sémantiques δ correspondent aux unités lexicales : par exemple $\delta 1$ est le hiéroglyphe sémantique de *sun*.
- les hiéroglyphes formels ϕ correspondent aux informations grammaticales : $\phi 5$ est le hiéroglyphe formel de l'article défini.
- les hiéroglyphes tectoniques traduisent l'ordre des mots : $\tau 4$ est le hiéroglyphe tectonique indiquant que l'article est suivi de N.

Cette langue se présente comme l'intersection des propriétés communes aux langues naturelles considérées et comme une langue univoque d'information. C'est ce qu'indique la référence au système hiéroglyphique qui souligne le caractère double du système de conventions permettant à la fois d'attribuer une signification sans ambiguïté aux signes primaires d'une part, et aux combinaisons d'autre part. Toutefois, comme on le verra dans le § 9 (ci-dessous), la question de la non-concordance des hiéroglyphes n'est pas sans poser quelques problèmes dans le processus de TA.

Langage de représentation sémantique abstrait

Comme nous l'avons évoqué plus haut (§ 5), des raisons liées à la nature propre des langues naturelles conduisent Mel'uk à l'idée d'une langue intermédiaire comme méthode de TA. En mettant au point un système de règles pour la traduction du hongrois en russe, il note l'intérêt pour la TA du hongrois qui présente des traits communs avec d'autres langues. Comme les langues turques, le hongrois est agglutinant. Comme l'allemand, il a des mots composés, des préfixes séparables, et des postpositions ayant la fonction de prépositions. Résoudre les problèmes de TA en hongrois permettrait donc de résoudre les problèmes de TA qui existent aussi dans d'autres langues.

Le hongrois présente toutefois une caractéristique spécifique : l'ordre des mots est radicalement différent de celui du russe. La solution consiste soit à concevoir une procédure spéciale qui rende compte de ce changement pour chaque paire de langues, soit à envisager une méthode dissociant analyse et synthèse de façon à neutraliser le problème de transfert de l'ordre des mots. Cette méthode, c'est la langue intermédiaire.

Sur le plan de la traduction, Mel'uk partage la position de son maître Ivanov, selon laquelle la traduction n'est pas directement liée au sens des mots et que seule une analyse sémantique peut rendre compte du transfert de sens entre une langue et une autre. Cette analyse doit être plus précise et plus objective que l'analyse sémantique traditionnelle : il faut réaliser un langage de représentation sémantique abstrait.

La langue intermédiaire de Mel'uk est, comme la proto-langue d'Ivanov (*cf.* Mel'uk 1959), fondée sur une théorie générale des relations entre systèmes linguistiques. Les langues sont des systèmes globaux, non séparables en niveaux, et dont une théorie sémantique doit rendre compte — conception que partageront également A.K. olkovskij (né en 1937) et Ju. D. Apresjan (né en 1930) qui ont mis au point avec Mel'uk le modèle sens-texte à partir de 1965.

La langue intermédiaire ne peut en aucun cas être une langue naturelle. C'est un système formel de correspondances entre langues naturelles qui n'a pas à être utilisé pour la communication : il ne se parle pas et n'a pas besoin de se matérialiser²⁶.

La langue intermédiaire est conçue comme un système abstrait formé de mots reliés par des relations syntaxiques. De ce point de vue, dit Mel'uk, elle est analogue à un langage formel logico-mathématique : les mots sont les symboles et les relations syntaxiques, les règles déductives.

On remarquera que, grâce à ces caractéristiques, ce langage s'apparente à une langue philosophique *a priori*²⁷. Cependant, contrairement aux langues universelles *a priori*, la langue intermédiaire de Mel'uk doit rendre compte des propriétés des langues naturelles, représenter leurs caractéristiques universelles tout en prenant en compte le problème spécifique de traduction d'une langue à l'autre. Mel'uk élabore donc un système sémantique dont les catégories sont des *significations* (*znaenija*) syntaxiques et lexicales. Les significations syntaxiques désignent les relations entre les éléments de la langue. Les significations lexicales désignent les liens entre la langue et le monde extérieur par rapport à la langue. Ces significations sont universelles : elles sont identiques pour toutes les langues ; elles doivent pouvoir être exprimées dans n'importe quelle langue ; enfin, tout énoncé d'une langue doit pouvoir être exprimé par ces significations. Toutefois, ce sont des abstractions qui n'ont pas nécessairement de réalisation concrète dans toutes les langues naturelles.

Le système abstrait de correspondances dépend du nombre de langues pour lequel il est élaboré. Si l'on ajoute de nouvelles langues, le système se trouve modifié. Il a donc un caractère cumulatif et évolutif et se présente comme l'union des propriétés des langues naturelles considérées, alors que le modèle d'Andreev se présentait comme leur intersection. Toutefois, bien que toutes les différences entre langues puissent être exprimées dans la langue intermédiaire, seules certaines, comme on va le voir, sont obligatoires. En effet, une des caractéristiques de la sémantique du langage abstrait est sa "pauvreté". Elle doit être réduite à un minimum d'ensembles sémantiques simples.

²⁶. Une langue intermédiaire sans matérialité linguistique est difficilement acceptable par les Soviétiques de l'époque. Ce sont les mathématiciens, tel Uspenskij (1959), qui encouragent le plus le projet de Mel'uk en reconnaissant derrière sa langue intermédiaire un projet de grammaire universelle.

²⁷. Les langues philosophiques *a priori* n'ont pas de finalité pratique de communication. Elles se présentent comme une grammaire des idées ou une classification des connaissances. Indépendantes des langues naturelles, elles sont postulées *a priori* : le travail du philosophe précède celui du linguiste (cf. Couturat et Léau 1903, *Histoire de la langue universelle*, Paris, Hachette).

9. LANGUES INTERMÉDIAIRES, VARIATIONS DES LANGUES ET TA

Dans ces modèles, la focalisation du processus de traduction sur le transfert de sens d'une langue à une autre en s'appuyant sur les invariants linguistiques se heurte précisément au problème de la variation. Autrement dit, que faire des traits variants dans la langue intermédiaire ?

Dans le modèle d'Andreev, c'est la non-concordance des hiéroglyphes entre langues source et cible qui pose le problème de la variation, et ceci à différents niveaux : (i) présence d'une caractéristique morpho-syntaxique dans une langue et pas dans l'autre (c'est, on le verra, le problème principal traité par Mel'uk) ; (ii) réalisation lexicale sur un nombre d'unités différentes qui implique un changement de types de hiéroglyphes. Ainsi *to catch cold* correspondant à *prostudit'sja* en russe entraîne le remplacement d'un hiéroglyphe sémantique par un hiéroglyphe formel ; (iii) la polysémie se traduit par une non-concordance sémantique.

Face aux énormes difficultés que posent, pour la TA, ces trois types de non-concordance, surtout quand elles se produisent simultanément, Andreev propose de supprimer de la langue intermédiaire les hiéroglyphes non concordants, tels que les articles et le genre dans les langues romanes et germaniques ou les classifieurs dans les langues bantoues. Cette solution a suscité de nombreuses critiques²⁸. Par ailleurs, étant donné les problèmes posés par l'automatisation de la traduction, le projet n'a pas donné lieu à des expériences en machine.

Pour Mel'uk (1960), les variations sont au cœur du dispositif. Toutes les significations grammaticales ne sont pas identiques au regard de la traduction. Une bonne partie des informations morpho-syntaxiques, singulières pour chaque langue et utiles lors de l'analyse du texte en langue source, ne sont plus nécessaires lors de la phase de traduction proprement dite. Ainsi il n'est pas nécessaire de connaître directement le genre, le nombre, le cas d'un adjectif russe quand on traduit un adjectif russe en français ou en allemand. Ce qui est important c'est d'avoir trouvé, lors de la phase d'analyse, le substantif avec lequel s'accorde l'adjectif.

Il en va autrement des indicateurs grammaticaux de significations lexicales, servant au repérage référentiel des éléments lexicaux (nombre des noms, temps et mode des verbes, etc.) qui, eux, sont nécessaires à la traduction.

Ainsi la langue intermédiaire comportera de façon obligatoire seulement deux sortes de variations entre langues, les significations lexicales et syn-

²⁸. Ce mode de construction fondé sur la congruence des langues, jugé intéressant *a priori*, a été vivement critiqué (cf. Lentin 1960, Uspenskij 1958) tant sur le plan mathématique (calcul de l'indice) que sur le plan linguistique : affectation aux langues d'un poids proportionnel et rejet de certains traits linguistiques sur la base de critères extralinguistiques, absence de hiérarchie des hiéroglyphes, notion de hiéroglyphe sémantique ne tenant pas compte de la polysémie....

taxiques²⁹ ; d'une part parce qu'elles existent dans toutes les langues : d'autre part parce qu'elles sont nécessaires à la traduction.

Ceci a des conséquences pour la structure de la langue intermédiaire qui, parce qu'elle ne doit pas rendre compte des significations grammaticales des différentes langues, n'a pas de vraie grammaire. En revanche, les significations syntaxiques sont stockées sous forme de relations syntaxiques de la langue intermédiaire, et les traits grammaticaux lexicaux sont stockés sous forme de mots (ou de concepts) dans un dictionnaire. Mel'uk rejoint ici la position d'Ivanov qu'il cite : " concrètement les significations relationnelles doivent se refléter non dans la grammaire, mais dans le vocabulaire de la machine " (1960).

Autrement dit, une partie considérable de la grammaire de la langue naturelle doit être traduite sous forme de dictionnaire. Ceci s'explique, dit-il, par le fait que la grammaire des langues naturelles correspond très peu à la syntaxe logique d'une langue scientifique. Les mots et les relations syntaxiques de la langue intermédiaire correspondent à l'alphabet et aux règles de formation des langages formels de la logique mathématique.

Enfin, le traitement automatique de la variation des langues impose deux contraintes lors de la traduction de la langue source vers la langue intermédiaire : (i) il ne faut perdre aucune information grammaticale de cette langue parce qu'une telle information peut être nécessaire, en synthèse, à la traduction dans une des langues cibles. Ainsi, la catégorie du nombre des noms sera exprimée dans la langue intermédiaire quand elle existe dans les langues sources (russe, anglais, hongrois...), mais non pour le chinois : ce qui évite à la fois perte et redondance de l'information. (ii) aucune information ne doit être développée de façon spécifique qui ne soit exprimée sous une forme claire dans la langue source, car cette information peut s'avérer superflue lors de la traduction dans la langue cible. Par exemple, lors de la traduction du japonais vers le chinois, l'introduction dans la langue intermédiaire du nombre des noms, qui n'est pas exprimé en japonais, exige la mise en œuvre d'une série de règles lexico-sémantiques, statistiques et contextuelles, inutile et redondante dans ce cas précis. Lorsque cette information est requise pour la langue de sortie, il est plus économique de la générer durant le processus de synthèse du texte, grâce à une analyse du contexte.

Une telle approche permet de traduire de nouvelles langues sans introduire de changements fondamentaux dans la langue intermédiaire : les nouvelles significations grammaticales sont transformées en mots de la langue intermédiaire et introduites de façon cumulative sans changer sa structure.

À la lumière de ces deux modèles, on peut dire que la notion de langue universelle se trouve modifiée par son appréhension comme langue intermé-

²⁹. L'opposition entre significations lexicales et significations grammaticales a pour origine la lecture d'un article de 1959 de R. Jakobson " Boas' view of grammatical meaning " – communication personnelle de I.A. Mel'uk.

diaire dans un système de TA. La nécessité de prendre en compte les caractéristiques communes des langues conduit à l'élaboration de formes mixtes, à la fois langue universelle a priori, langue logique ou langue univoque d'information, et a posteriori puisque rendant compte des caractéristiques des langues naturelles. L'automatisation de la traduction a pour effet supplémentaire de la rendre dynamique, du moins chez Mel'uk, où les ingrédients de la langue intermédiaire ne sont pas fixés d'avance mais dépendent de la paire de langues considérées.

CONCLUSION

Pour conclure, il est intéressant de noter que l'élaboration d'une langue intermédiaire a plusieurs conséquences. La TA n'est plus seulement un objectif pratique, c'est le lieu où peut être pensé le rapport entre les langues et la recherche d'universaux de façon systématique. Reste à savoir le devenir qu'a eu cette systématisation sur les recherches en linguistique en URSS. S'est-elle investie dans une réelle démarche de typologie expérimentale des langues à laquelle elle semble apparentée ?

La mise en algorithme des caractéristiques communes aux langues dans la langue intermédiaire a pour conséquence de ne pas considérer les flexions (terme qu'utilisera Mel'uk dans son cours de morphologie générale pour désigner les significations grammaticales) comme des universaux. Marquant les singularités des langues, elles ne peuvent être utilisées que pour la phase d'analyse dans un système de TA. Toutefois, pour Mel'uk, c'est un véritable fait linguistique et non simplement une nécessité imposée par l'algorithme.

À cet égard, il est intéressant de noter que la position des Soviétiques concernant le statut de la linguistique dans la TA, position particulièrement illustrée par Mel'uk, est radicalement opposée à celle de certains expérimentateurs américains. Ces derniers revendiquent une position spécifique, indépendante des linguistes (*cf.* Reifler in Locke et Booth 1955) et se donnent des objectifs exclusivement pratiques, abandonnant explicitement la connaissance de la structure des langues aux linguistes théoriciens.

Très différente est la position de Mel'uk. La construction d'une langue intermédiaire est un travail de linguiste. Il va même, dans le recueil sur la TA publié après le rapport de l'ALPAC (*cf.* Kulagina, Mel'uk 1967), jusqu'à tenir l'échec de la TA pour un échec de la linguistique. Tout progrès en TA, dit-il, est subordonné aux progrès de la linguistique qui doit résoudre deux tâches : créer un langage de représentation sémantique et compiler des algorithmes d'analyse et de synthèse de textes : cela n'étant possible que si la linguistique est transformée en connexion avec les mathématiques. On voit que la naissance de la linguistique mathématique³⁰ en URSS est fondée sur des bases

³⁰0. En URSS, le terme *linguistique mathématique* correspond en partie à ce qui aux Etats-Unis est appelé linguistique computationnelle, à l'intersection de la linguistique

radicalement différentes de celles de la linguistique computationnelle aux États-Unis (*cf.* Léon 1992, 1996). C'est une discipline à part entière, appartenant à la linguistique.

On peut considérer que les tentatives de construction d'un tel modèle constituent toutes, même l'entreprise du groupe de Panov, un pas vers la formalisation des langues³¹, étape incontournable du traitement automatique des langues. Elles permettent d'échapper à la méthode du mot à mot qui, en imposant à l'analyse linguistique de se conformer directement à l'organisation d'une machine électronique, ne peut aboutir qu'à des impasses sur le plan linguistique comme sur le plan du processus de traduction. La langue intermédiaire³² sera à l'origine d'un certain nombre de développements linguistiques originaux, notamment par les linguistes Apresjan, olkovskij et Mel'uk, de la syntaxe de dépendance et des fonctions lexicales.

reçu juin 1997

adresse des auteurs :
 Université Paris VII
 URA CNRS 381
 Tour Centrale - Bureau 818
 2, place Jussieu
 75251 - Paris Cedex 05

proprement dite, des langages formels offerts par la logique mathématique et certaines contraintes inhérentes au traitement automatique de l'information. Elle inclut donc la théorie des automates et la théorie des algorithmes. Toutefois la linguistique mathématique est susceptible de développer de façon purement théorique, indépendamment de l'existence même de machines concrètes.

- ³¹. On peut considérer à juste titre ces modèles de langue intermédiaire comme les premiers langages de représentation sémantique qui connurent un développement considérable dans le domaine de l'intelligence artificielle à partir des années 1980. Il ne faut pas oublier qu'ils ont eu également une postérité dans les recherches ultérieures en TA : le langage pivot développé par le GETA au début des années 60 s'inspire, pour la syntaxe, du modèle de Tesnière et pour la sémantique lexicale du modèle de Mel'uk. C'est avant tout un modèle syntaxique, fondé sur des arbres syntaxiques profonds "très sémantisés".
- ³². Dans le modèle sens-texte, la langue intermédiaire dont la conception restait attachée à la réalisation d'algorithmes concrets, a été remplacée par l'anglais *basic* et le russe *basic*, ensemble de structures lexico-syntaxiques profondes, nécessaire à la synthèse. Pour la TA, cette approche reste liée aux exigences de la demande sociale puisqu'elle permet à un utilisateur ne connaissant que l'anglais de traduire un texte de l'anglais vers le russe.

RÉFÉRENCES

- ANDREEV, N.D. (1957). "Mainnyj perevod : problema jazyka-posrednika", *Voprosy Jazykoznanija* 5, 117-121 [Trad. angl. : "Machine Translation and the problem of an Intermediary Language", JPRS/68, archives ATALA].
- ANDREEV, N.D., IVANOV, V.V., MEL'UK, I.A. (1960). "Nekotorye zameanija i predloenija otositel'no raboty po maöinnomu perevodu v SSSR", *Mainnyj Perevod i Prikladnaja Lingvistika* 4, 3-24 [Trad. angl. : "Some remarks and suggestions to work on machine translation in the USSR", JPRS/8026, archives ATALA].
- AUROUX, S. (1979). *La Sémiotique des encyclopédistes. Essai d'épistémologie historique des sciences du langage*, Paris, Payot.
- BEL'SKAJA, I.K. (1958). "Osnovnye karakteristiki slovarja i grammatičeskix sxem maöinnogo perevoda s anglijskogo jazyka na russkij", *Sbornik statej po maöinnomu perevodu*, 47-80, Moskva [Trad. angl. : "Basic Characteristics of the Dictionary and Grammatical Routines for MT from English into Russian", JPRS/925, archives ATALA].
- BODUEN DE KURTENE, I. (1871). "Obie zameanija o jazykovedenii i o jazyke [Remarques générales sur la linguistique et sur le langage]", *urnal Ministerstva Narodnago Prosveenija*, Tom CLVIII -1871, 279-316.
- BODUEN DE KURTENE, I. (1901). "O smeannom karaktere vsex jazykov [Du caractère mixte de toutes les langues]", *urnal Ministerstva Narodnago Prosveenija*, Tom CCCXXXVII, 12-24.
- CAUSSAT, P., ADAMSKI, D., CREPON, M. (1996). *La Langue, source de la nation ; messianismes séculiers en Europe Centrale et Orientale du XVIII^e siècle au XX^e siècle*, Liège, Pierre Mardaga.
- CEJTIN, G.S. (1997). "Let's rewrite the Rules of the Road", *The Newsletter of the European Network in Language and Speech* 6.1.
- DREZEN, E.K. (1936). *Internacionalisacija naučno-terminološkoj terminologii, Istorija, sovremennoe položenie i perspektivy*, Moskva-Leningrad, Gosudarstvennoje izdatel'stvo po standardisacii "Standartgiz". [Internationalisation de la terminologie scientifique et technique, Histoire, actualité et perspectives].
- HARPER, K. (1960). "Soviet research in machine translation", *Proceedings of the National Symposium on Machine Translation*, 2-12.
- HUTCHINS, W.J. (1986). *Machine Translation, past, present, future*, Chichester, Ellis Horwood ltd.
- HUTCHINS, W.J. (1993). "The first MT patents", *MT News International* 5, 14-15.
- IVANOV, V.V. (1958). "Tipologija i sravnitel'no-istoričeskoje jazykoznanie [Typologie et linguistique historique et comparée]", *Voprosy Jazykoznanija* 5, 34-39.
- IVANOV, V.V. (1959). "Théorie de la compatibilité entre les systèmes de langage et les bases de la linguistique historico-comparative", *Extraits des rapports présentés à la Conférence de Linguistique Mathématique*, 5-11, Leningrad, Éditions du ministère de l'Instruction Supérieure [Trad. fr. Comité d'Action Scientifique de Défense Nationale, Centre d'exploitation du renseignement scientifique et technique, archives ATALA].
- KULAGINA, O.S., MEL'UK, I.A. (1956). "O maöinnom perevode s francuzskogo na russkij", *Voprosy Jazykoznanija* 5, 111-121 [Trad. angl. : "Machine Translation from French to Russian", JPRS/DC-68, archives ATALA].
- KULAGINA, O.S., MEL'UK, I.A. (1967). "Automatic translation : some theoretical aspects and the design of a translation system", Booth, A.D. (éd.), *Machine Translation*, 137-173, Amsterdam, North Holland Publishing Company.

- KUZNECOV, P.S., LJAPUNOV, A.A., REFORMACKIY, A.A. (1956). "Osnovnye problemy maõinnogo perevoda", *Voprosy Jazykoznanija* 5, 107-111 [Trad. angl. : "Fundamental problems of machine translation", JPRS/68, archives ATALA].
- KUZNECOV, S. (1995). "Linguistica cosmica : la naissance du paradigme cosmique", *Histoire Epistémologie Langage* XVII/2, 211-234.
- LENTIN, A. (1960). "La Notion de congruence", *La Traduction Automatique* 2, 11-17.
- LEON, J. (1992). "De la traduction automatique à la linguistique computationnelle. Contribution à une chronologie des années 1959-1965", *Traitement Automatique des Langues*, N° spécial trentenaire, vol. 33, n° 1-2, 25-44.
- LEON, J. (à paraître). "La Traduction automatique I : les premières tentatives jusqu'au rapport ALPAC", *Handbücher zur Sprach- und Kommunikationswissenschaft*, Berlin, Walter de Gruyter and co., vol. 3, *Histoire des Sciences du Langage*.
- Lingvistijskij enciklopedijskij slovar'* (1990), Moskva, Sovetskaja Enciklopedija.
- LOCKE, W.N., BOOTH, A.D. (éds) (1955). *Machine Translation of Languages*, 14 essays, New York, MIT et John Wiley.
- MEL'UK, I.A. (1958). "O maõinnom perevode s vengerskogo na russkij", *Problemy Kibernetiki* 1, 222-264 [Trad. angl. : "Translation from Hungarian into Russian", JPRS/646, archives ATALA].
- MEL'UK, I.A. (1959). "Raboty po maõinnomu perevodu v SSSR", *Vestnik Akademii Nauk SSSR* 2, 43-47, [Trad. angl. : "Work on Machine Translation in the USSR", JPRS/662, archives ATALA].
- MEL'UK, I.A. (1960). "K voprosu o grammatičeskom v jazyke-posrednike", *Mainnyj Perevod i Prikladnaja Linguistika* 4, 25-451 [Trad. angl. : "The problem concerning the "grammatical", "in an intermediate language", JPRS/8026, archives ATALA].
- MEL'UK, I.A. (1961). "Nekotorye voprosy maõinnogo perevoda za rubežom", *Institut Naučnoj Informacii* 6, 1-44, Moskva, AN SSSR [Trad. angl. : "Some problems of MT abroad, USSR", *Reports at the conference on Information Processing, MT and Automatic Text Reading*, JPRS/13135, archives ATALA].
- MEL'UK, I.A., RAVI, R.D. (1967). *Avtomatičeskij perevod (1949-1963)*, Moskva, AN SSSR.
- MOLONAJA, T.N. (1957). "Nekotorye voprosy sintaksisa v svjazi s maõinnym perevodom s anglijskogo jazyka na russkij", *Voprosy Jazykoznanija* 4, 46-59 [Trad. angl. : "Certain questions of syntax in connection with machine translation from English to Russian", JPRS/68, archives ATALA].
- NIKOLAËVA, T.M. (1958). "Konferencija po maõinnomu perevodu", *Voprosy Jazykoznanija* 5, 149-151, [Trad. angl. : "Conference on MT : compte rendu du 1er congrès sur la TA, Moscou, 15-21 mai 58", JPRS/487, archives ATALA].
- PANOV, D.Ju. (1956). *Avtomatičeskij perevod*, Moskva, AN SSSR [Trad. angl. : 1960, *Automatic Translation*, Oxford, New York, Pergamon Press Inc].
- RASMUSSEN, A. (1995). *L'Internationale scientifique (1890-1914)*, Thèse de doctorat d'histoire, Paris, EHESS.
- SALMON, V. (1992). "Caractéristiques et langues universelles", Auroux, S. (éd.), *Histoire des Idées Linguistiques*, t. II, 407-423, Liège, Mardaga.
- TATARINOV, V.A. (1993). "Iz istorii oteestvennogo terminovedenija : Ernest Karlovi Drezén", *Voprosy Jazykoznanija* 3, 113-119 [Histoire de la terminologie en Russie : E.K. Drezén].

USPENSKIJ, V.A. (1959). "Itogi raboty sekcii algoritmov maöinnogo perevoda", Moscou, *Mainnyj Perevod i Prikladnaya Lingvistika* 1, 31-62 [Trad. angl. : "Results of the work of the MT algorithm section", Actes du 1er congrès de TA, 15-21 mai 1958, JPRS/1006, archives ATALA].