Meta
Journal des traducteurs
Translators' Journal

META

Le système TEAM, une aide à la traduction

Joachim Schulz

Volume 16, numéro 1-2, mars 1971

Actes du colloque international de linguistique et de traduction.
Montréal, 30 septembre - 3 octobre 1970

URI : https://id.erudit.org/iderudit/004055ar
DOI : https://doi.org/10.7202/004055ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Schulz, J. (1971). Le système TEAM, une aide à la traduction. *Meta*, *16*(1-2), 125–131. https://doi.org/10.7202/004055ar

Le système TEAM, une aide à la traduction

L'explosion des connaissances que l'on observe dans le monde entier, l'expansion des industries et des marchés au-delà des frontières nationales sont à l'origine d'un échange d'informations qui ne cesse de s'intensifier. Cette tendance à l'internationalisation des rapports de tous genres pose évidemment le problème très délicat de la traduction des documents relevant des domaines les plus divers. Et comme partout où l'on se heurte à des questions de grande complexité, on fait appel à l'informatique. L'ordinateur est mis à contribution pour résoudre des problèmes linguistiques.

Le besoin d'une aide à la traduction efficace se fait naturellement tout particulièrement sentir dans une firme internationale comme Siemens dont les activités s'étendent au monde entier et qui est confrontée jour après jour à un flux permanent d'informations commerciales, techniques et scientifiques qu'il s'agit de traduire rapidement et correctement dans les langues les plus diverses.

Le problème central est évidemment celui de la terminologie : l'évolution vertigineuse des sciences et des techniques donne constamment naissance à de nouveaux termes, de nouvelles expressions. Le traducteur, même spécialisé, n'est plus en mesure de lire, ne serait-ce qu'en diagonale, toute la littérature intéressant son domaine de travail et encore moins de dépouiller les textes en vue d'établir un fichier. La recherche d'un terme technique inconnu à l'aide des moyens conventionnels est souvent tellement longue qu'il a bien fallu s'écarter des sentiers battus de la lexicographie et trouver des solutions inédites à des problèmes nouveaux.

C'est la raison pour laquelle le service de traduction de la société Siemens a mis au point un ensemble de programmes que nous avons baptisé TEAM. T-E-A-M, ce qui veut dire en allemand *Terminologie-Erfassungs- und Auswertungs-Methode,* méthode destinée à la saisie et à l'exploitation des terminologies.

L'élément central en est un dictionnaire électronique, ou comme dirait l'informaticien, un dictionnaire mis en mémoire dans un ordinateur. Ce dictionnaire stocké sur bande magnétique est une sorte de répertoire multilingue de termes techniques, qui est constamment mis à jour. Il ne s'agit donc pas, et j'insiste sur ce fait, d'une traduction automatique par ordinateur, mais tout simplement d'une aide à la traduction qui met à la disposition du traducteur la terminologie technique la plus récente mémorisée dans un ensemble électronique.

Quelles sont donc les informations qui sont stockées dans ce dictionnaire et comment l'utilisateur peut-il y accéder ? Voilà deux problèmes que je me propose de développer plus en détail.

Chaque unité du dictionnaire électronique, c'est-à-dire chaque enregistrement, contient, dans les différentes langues utilisées, tous les termes relatifs à une notion donnée, y compris les synonymes. Il s'y ajoute, le cas échéant, des informations supplémentaires telles que le genre des mots, verbe ou substantif (masculin ou féminin), la source, c'est-à-dire le lieu où le terme a été trouvé, le ou les secteurs techniques d'application, un critère de qualité, la date d'entrée, des définitions, des exemples de contexte, etc. À chacun de ces enregistrements on affecte un numéro qui permet de trouver et d'adresser de façon absolument univoque, la rubrique recherchée.

Les différentes catégories d'informations de ces rubriques peuvent avoir une longueur quelconque. Il est possible de les adresser séparément au sein d'un enregistrement, ce qui revient à dire qu'elles sont faciles à corriger, à compléter ou à effacer.

L'indicatif du domaine technique dont relèvent les différents termes est extrêmement important pour le traducteur. Il aide à résoudre le problème des homonymes, et renvoie le traducteur au secteur technique précis où le terme est appliqué. Dans notre système, nous utilisons, pour caractériser les différents secteurs techniques, un code mis au point par l'Union de documentation de l'industrie électrique allemande. Il s'agit d'une sorte de clef décimale que l'on peut subdiviser à volonté en ajoutant des décimales. Puisqu'il est possible d'affecter plusieurs de ces clefs à chaque terme, on obtient une classification suffisamment souple et explicite.

L'introduction des enregistrements, qu'il s'agisse de termes, de corrections ou de mises à jour se fait par ruban perforé. Les téléimprimeurs utilisés à cet effet permettent de respecter rigoureusement l'orthographe c'est-à-dire l'écriture de lettres majuscules et minuscules ainsi que des signes diacritiques les plus divers tels que accents, cédilles, etc., les voyelles infléchies des langues germaniques et ainsi de suite. Étant donné que nous enregistrons également le russe et que le clavier du téléimprimeur ne dispose que de l'alphabet latin, les lettres cyrilliques sont restituées sous forme translittérée. Cette translittération qui repose sur l'alphabet tchèque est réalisée selon l'avis ISO R9. Elle ne pose aucun problème ni au niveau de la lecture, ni à celui de la correction et permet, en outre, la retranslittération exacte, et par conséquent l'édition d'un dictionnaire en lettres cyrilliques.

J'aimerais approfondir encore un peu le problème de la translittération ou plus précisément le problème de la réversibilité de la translittération. À la différence de la transcription qui, elle, s'occupe de la restitution de valeurs phoniques, la translittération ne fait que substituer aux lettres d'un alphabet, en l'occurrence l'alphabet cyrillique, les lettres d'un autre alphabet, dans notre cas, l'alphabet latin. Le fait que le nombre des lettres cyrilliques est supérieur à celui des lettres latines crée certaines difficultés. Ainsi, à chaque lettre cyrillique doit rigoureusement correspondre une lettre latine, ou une combinaison déterminée de lettres latines et de signes diacritiques, et *vice versa,* c'est-à-dire que la correspondance

doit être biunivoque comme on dit en jargon informatique : à chaque lettre ou à chaque combinaison de lettres d'un mot translittéré ne doit correspondre qu'une seule lettre cyrillique. Dans le cas d'une combinaison de lettres, cette biunivocité n'est pas assurée du point de vue logico-mathématique pur : ainsi la lettre russe «ю» est représentée par deux lettres latines (*j* + *u*) qui, prises séparément, correspondent cependant à deux autres lettres russes (ĭy) . Dans une retranslittération automatique, la combinaison *j u* donnerait aussi bien la lettre «ю» que les deux caractères «ĭy» . Mais puisqu'il ne s'agit pas de combinaisons logiques purement abstraites, mais d'un problème linguistique, à savoir l'apparition de lettres isolées au sein d'un mot, il est facile de tourner la difficulté. La combinaison «ĭy» n'existe pas dans le russe moderne. La valeur phonique correspondant à cette combinaison est toujours exprimée par le caractère spécial «ю». La combinaison latine *j u* donne donc forcément la lettre cyrillique «ю». On résoud de façon analogue les quelques autres problèmes de retranslittération complexes. Il s'avère donc que le schéma ISO peut être utilisé pour programmer, à l'aide de l'ordinateur, les règles linguistiques régissant la translittération et la retranslittération de l'alphabet russe.

Nous venons maintenant au problème du traitement automatique du dictionnaire électronique. Comment les informations enregistrées en mémoire sont-elles mises à la disposition de l'utiliseur ? Je me bornerai à la description de deux possibilités fondamentales : *primo,* l'interrogation directe du fichier électronique à l'aide d'un écran cathodique, et *secundo* l'édition de dictionnaires conventionnels sous forme imprimée.

Permettez-moi d'aborder d'abord quelques questions d'ordre général liées spécialement à l'utilisation d'un dictionnaire électronique ou conventionnel :

La première question concerne les synonymes. Nous avons déjà dit que chacune des rubriques enregistrées peut également contenir des synonymes, et ce dans toutes les langues désirées. L'utilisateur d'un dictionnaire ou bien d'un système électronique ne veut pas seulement trouver dans la langue cible tous les synonymes relatifs à un terme donné. Il veut également trouver ces informations sous chacun des synonymes existants dans la langue source. Afin de pouvoir satisfaire à cette exigence, on génère automatiquement, c'est-à-dire par programme, des rubriques spéciales pour tous les synonymes dans la langue source d'un dictionnaire.

La deuxième question concerne les termes proprements dits. Il faut bien souligner qu'un terme peut consister en un seul mot ou en comprendre plusieurs. Les termes composés de plusieurs mots sont très fréquents dans le langage technique. Ainsi, au terme allemand *Datenverarbeitungsanlage* correspond en français l'expression « ensemble électronique de traitement de l'information » qui se compose de trois substantifs et d'un adjectif. Ces compléments déterminatifs ou d'appartenance sont très fréquents en français. En anglais plusieurs substantifs formant un terme sont tout simplement juxtaposés et en allemand on rencontre très souvent des termes composés d'un substantif précédé d'un adjectif.

Dans tous ces cas on peut trouver l'information recherchée non seulement sous le premier mot d'une expression mais aussi sous tous les autres, à condition que ceux-ci aient suffisamment de poids. Ces mots de poids supérieur ou autrement dit les mots clefs d'un terme composé sont affectés d'un signe de commande

lors de l'entrée des données. À l'aide de ces signes de commande, le programme génère des inversions, c'est-à-dire des rubriques où le mot clef apparaît en première position alors que tout le reste suit, séparé du mot clef par une virgule. Ces rubriques dites d'inversion peuvent comprendre toutes les informations de la rubrique primitive ou bien tout simplement contenir un renvoi à cette rubrique.

De la même façon, on traite les abréviations très courantes dans les terminologies techniques, car on veut trouver dans le dictionnaire les deux formes d'un terme, en abrégé et en toutes lettres.

Un autre problème que pose la consultation d'un dictionnaire électronique relève de certaines incohérences de l'orthographe comme on les rencontre tout particulièrement en anglais. Ainsi écrit-on en anglais les termes composés tantôt avec un trait d'union tantôt sans trait d'union et parfois même en un seul mot. Il n'existe pas de règle, les grammaires restent muettes sur ce point. Or, la machine ne trouve la réponse à une question donnée qu'en comparant caractère par caractère le terme recherché avec les termes enregistrés en mémoire. S'il y a donc une différence tant soit peu légère entre l'orthographe du terme recherché et celle du terme enregistré, la machine n'est pas capable de sortir une réponse, même si elle était théoriquement en mesure de la donner. Nous évitons cet inconvénient en ramenant automatiquement tous les termes à une forme normale déterminée, c'est-à-dire en supprimant non seulement tous les blancs et traits d'union, mais aussi les accents et les signes d'interponctuation ainsi que la différence entre les majuscules et les minuscules. Cette forme normale qu'on appelle également argument de tri permet un classement exact des termes par ordre alphabétique. C'est grâce à elle que la machine est à même de reconnaître l'identité de termes qui ne se différencient les uns des autres que par cette orthographe fluctuante dont je viens de vous parler.

La réduction de tous les termes à une forme normale est très importante pour le contrôle interne, lorsqu'on vérifie si un nouveau terme qu'on vient d'introduire n'est pas déjà enregistré en mémoire. Si tel est le cas, la rubrique nouvellement introduite est supprimée automatiquement à l'exception des informations supplémentaires qui l'accompagnent et qui n'existaient pas encore dans l'enregistrement primitif. Mais la forme normale ne sert pas seulement à résoudre le problème des doublets. Elle est une condition préalable à toute interrogation du dictionnaire électronique par l'homme.

Il existe bien sûr d'autres problèmes d'orthographe fluctuante dont la machine ne vient pas à bout au moyen de la forme normale. Il suffit de mentionner les différences entre l'orthographe anglaise et l'orthographe américaine, par exemple *synchronise* avec un *z* ou avec un *s*. Dans ce cas, on peut soit enregistrer les deux formes en tant que synonymes, ou bien l'utilisateur doit éventuellement chercher le terme sous les deux formes. Pour terminer ce chapitre sur l'orthographe, il faut bien avouer, que contrairement à l'homme, la machine n'est pas en mesure de reconnaître les fautes de frappe ou autres erreurs. Si un mot n'est pas écrit correctement, la machine le qualifie d'inconnu et ne fournit pas de réponse.

Comme je l'ai déjà dit, l'utilisateur, pour obtenir les informations désirées, peut engager une sorte de dialogue avec la machine par l'intermédiaire, par exemple, d'un terminal à écran. Pour cette consultation directe du dictionnaire

électronique le TEAM est combiné avec un *Information Retrieval System,* un système de recouvrement de l'information, appelé « Golem ». La condition *sine qua non* d'un tel système conversationnel est une mémoire de masse à accès direct. Nous utilisons une mémoire à cartes magnétiques qui est adressée par l'intermédiaire de disques magnétiques. On retrouve les informations mises en mémoire à l'aide de descripteurs. Dans notre système, ce sont les termes eux-mêmes et le code-matière qui servent de descripteurs. L'utilisateur a la possibilité de combiner plusieurs descripteurs par les fonctions logiques ET, OU et NON, c'est-à-dire qu'il peut se faire sortir tous les équivalents d'un terme et de ses synonymes ou bien seulement ceux des équivalents qui relèvent d'un secteur technique déterminé, ou ceux qui n'en relèvent pas.

Et voilà pour l'aspect théorique du problème. Venons-en à la pratique : la situation de départ est évidemment classique. Il y a un texte à traduire qui contient des termes que le traducteur ne connaît pas. Alors que fait-il ? Au lieu de consulter un dictionnaire imprimé, le traducteur frappe au clavier de son terminal les termes dont il désire recevoir la traduction en précisant leur langue d'origine. La question posée est en même temps visualisée sur l'écran ce qui permet, le cas échéant, de la corriger ou de la compléter. La réponse donnée immédiatement par l'ordinateur apparaît également sur l'écran. L'information fournie par la machine peut avoir une longueur quelconque. Si elle déborde la capacité d'écriture de l'écran, elle est sortie en plusieurs séquences. Bien entendu, l'utilisateur peut poser plusieurs questions en même temps. Il reçoit alors les réponses automatiquement les unes après les autres.

Outre le dictionnaire, le « Golem » permet de mettre à la disposition du traducteur d'autres pools d'information encore, par exemple, des renseignements techniques sur un sujet donné.

Le terminal à écran utilisé actuellement n'autorise pour la visualisation des informations mises en mémoire qu'une orthographe simplifiée, c'est-à-dire que le texte est inscrit en lettres latines majuscules sans accents, etc. Le nombre des terminaux pouvant être reliés simultanément au système se trouve, du moins pour le moment, limité par certaines contraintes techniques.

Pour cette raison et d'autres encore, les dictionnaires conventionnels ne disparaîtront pas de sitôt de la table du traducteur. La mise en œuvre de moyens techniques modernes apportera, cependant, des améliorations importantes en matière d'aide à la traduction classique. Déjà, il est possible de publier, dans des délais très brefs, des dictionnaires techniques contenant la terminologie la plus récente. Ainsi la société Siemens vient-elle de sortir un dictionnaire de l'informatique dont la confection, de la clôture de la rédaction au livre imprimé en passant par la composition du texte, n'a pris que quatre semaines.

La fabrication d'un dictionnaire conventionnel imprimé à partir du dictionnaire électronique continuellement mis à jour comprend trois phases essentielles commandées par programme, donc effectuées automatiquement. Ces trois phases sont : la sélection d'une certaine quantité de termes à partir du répertoire global ; le tri des données, c'est-à-dire le classement des termes par ordre alphabétique ; et, finalement, la confection automatique d'un support d'impression, ou autrement dit la composition du texte page par page sur un film.

Mais revenons à la première phase, c'est-à-dire à la sélection : les dictionnaires techniques comprennent, en règle générale, la terminologie de domaines d'activités bien définis. On choisit les termes en fonction des clefs-matières affectées, comme nous l'avons déjà dit, à tous les enregistrements. Lorsque le dictionnaire est censé contenir plusieurs domaines voisins, on sélectionne toutes les notions qui appartiennent au moins à l'un des domaines en question (il s'agit dans ce cas, de l'opération logique OU). Mais on pourrait également limiter la sélection à ceux des termes que l'on retrouve dans tous les domaines considérés (il s'agit, dans ce cas de l'opération logique ET).

Puisqu'on stocke dans le dictionnaire électronique, outre les termes consacrés, également des termes de travail provisoires, la sélection peut aussi se faire selon la qualité des enregistrements. Il arrive souvent qu'on dispose d'un terme précis dans la langue source, auquel ne correspond dans la langue cible qu'une sorte de paraphrase. Celle-ci, si elle est utile au traducteur, ne peut évidemment pas figurer dans le dictionnaire en tant que terme source. Ce problème est également résolu par programme à l'aide d'un caractère de commande.

Chaque dictionnaire ne comprendra qu'un nombre déterminé de langues. Il faut donc opérer une sélection pour obtenir celles des rubriques qui contiennent les informations dans les langues désirées. Mais on peut également extraire certaines catégories d'informations des rubriques elles-mêmes. On détermine, à l'aide de paramètres de programme, les informations que l'on veut ajouter aux termes proprement dits comme, par exemple, source, définitions, etc.

En même temps que la sélection, on réalise une autre phase : on constitue, pour la langue source, l'argument de tri qui permet le classement correct des termes selon les règles régissant les langues et les alphabets en question. Ainsi, on néglige, lors du tri des termes français, les accents. En allemand, on classe les voyelles infléchies sous les voyelles simples. En espagnol, on traite les lettres *ñ* et *ll* comme les lettres autonomes. Il va sans dire que les termes russes sont classés suivant les règles de l'alphabet cyrillique russe, sans tenir compte de la translittération en lettres latines dont nous avons déjà parlé. Les chiffres arabes et romains, les signes d'interponctuation, les traits d'union, etc., sont toujours négligés lors du tri, et ce dans toutes les langues.

La troisième phase qu'implique la confection automatique d'un dictionnaire imprimé est la création d'un support d'impression. Le moyen technique le plus moderne dont nous disposons aujourd'hui en ce domaine, est la photocomposeuse électronique. Nous utilisons le système « Digiset » de la société Hell de Kiel. La commande de ces photocomposeuses est avantageusement assurée par des bandes magnétiques qui, outre le texte à imprimer, contiennent toutes les instructions typographiques. Les différentes lettres sont visualisées sur l'écran d'un tube cathodique en fonction de ces instructions, et projetées de là sur un film.

Le programme utilisé pour l'établissement de cette bande magnétique d'entrée assume plusieurs fonctions : pour l'impression des différentes catégories d'informations d'une rubrique on emploie différents styles de caractères. Ainsi les mots de la langue de départ sont imprimés en caractères demi-gras, et ceux de la langue cible, en caractères maigres. Les indications relatives au genre des mots sont mises en italique. Les lettres pourvues d'un accent posent un problème spécial :

on utilise, la plupart du temps, des accents dits flottants, c'est-à-dire des accents qui ne font pas partie intégrante de la lettre. Le programme doit calculer le positionnement exact de l'accent sur la lettre en fonction de la chasse de l'accent et de la lettre en question ainsi que du corps de la police utilisée. En construisant les lignes, le programme doit, en outre, vérifier si le texte ne déborde pas la ligne. Dans l'affirmative, le dernier mot serait rejeté à la ligne suivante. Nous n'effectuons pas la division syllabique automatique, du fait que nous ne disposons pas encore des programmes nécessaires pour toutes les langues de notre dictionnaire. Le texte est donc composé en drapeau. Les différentes lignes, ou plus précisément demi-lignes, doivent alors être aménagées dans les sens horizontal et vertical de façon à obtenir une page de dictionnaire à deux colonnes. Les initiales sont, lorsqu'elles apparaissent pour la première fois dans la langue source, imprimées en un corps plus grand et la colonne est interrompue. Dans un dictionnaire bilingue, on peut grouper sous un mot clef tous les termes composés contenant ce mot clef, qui à la répétition est remplacé par un tilde. On imprime d'abord les termes qui commencent par le mot clef, puis ceux où le mot clef ne se trouve pas en première position. Tous les termes groupés sous un mot clef sont à leur tour classés dans un ordre strictement alphabétique. Finalement, chaque page est foliotée et affectée d'un titre courant. Sur les pages impaires, le numéro est placé sur la gauche et le titre sur la droite, sur les pages paires, c'est l'inverse. Toutes les opérations que je viens de décrire sont effectuées de façon entièrement automatique.

J'espère que cet exposé sur notre système TEAM vous a permis de vous faire une idée de l'aide que la machine peut apporter à l'homme en matière de traduction de textes techniques.

Je ne voudrais pas terminer mon tour d'horizon sans jeter un regard sur l'évolution future de l'aide automatique à la traduction. Deux tendances fondamentales semblent se dessiner parmi beaucoup d'autres : la première va dans le sens d'un perfectionnement des moyens techniques. La consultation à distance d'un dictionnaire électronique sera de plus en plus facilitée, le volume des données mises en mémoire, éventuellement non plus sous forme numérique mais sous forme analogue par mémorisation d'images, augmentera sans cesse et l'utilisateur pourra y accéder dans les délais les plus brefs, à tout moment, et où qu'il se trouve. La deuxième tendance va dans le sens d'une suppression pure et simple de la consultation de la machine par l'homme : un texte mis sous une forme exploitable par l'ensemble électronique, est entré dans l'ordinateur. La machine y relève automatiquement les termes techniques et fournit une liste des mots traduits, et ceci dans l'ordre de leur apparition dans le texte.

JOACHIM SCHULZ