

# **Le rôle d'une banque de terminologie au service de la traduction**

par

**R. GOFFIN**

*Université Libre de Bruxelles - BELGIQUE  
Commission des Communautés Européennes*

435

## I.- Les banques de terminologie multilingue

Les procédures d'obédience informatique ont été mises au service de la lexicologie dès 1949. L'ordinateur a d'abord été une machine de stockage, de tri et de dépistage, un moyen de dresser des listes alphabétiques, des index de fréquences ou des tables de concordances. La mise sur ordinateur de l'oeuvre de Saint Thomas d'Aquin a commencé en 1949 par le Père R. Busa. La lexicologie automatique était bien partie.

Les premières banques de données linguistiques ont été les banques portant sur le domaine lexical. On projette aujourd'hui des banques de données textuelles et des banques de données orthographiques et grammaticales.

Depuis la fin des années 60, la recherche terminologique a partie liée avec l'informatique. Le système LEXIS de l'Office fédéral des langues à Hürth, le système TEAM (Terminologie - Erfassungs- und Auswertungsmethode) de la société Siemens à Munich, la Banque de terminologie TERMIUM de l'Université de Montréal, passée depuis 1975 au Bureau des traductions d'Ottawa, la Banque de terminologie du Québec (BTQ), conçue et développée par l'Office de la langue française du Gouvernement du Québec, la banque NORMATERM constituée par l'Association française de normalisation (AFNOS) à Paris et la Banque de données terminologiques EURODICAUTOM, conçue et développée par la Commission des Communautés européennes à Bruxelles et à Luxembourg : toutes ces aides électroniques prouvent à suffisance que la terminologie automatisée se porte bien.

Les cinq banques existantes visent toutes à rassembler, centraliser et diffuser de façon sélective des terminologies unilingues ou multilingues. Elles ont été conçues pour répondre à des besoins de communication scientifique et technique et servir une clientèle spécifique : traducteurs, rédacteurs, enseignants, le cas échéant, le grand public.

La BTQ s'adresse à un clientèle essentiellement francophone en quête d'équivalents bilingues (anglais-français), NORMATERM engrange et rend disponible la terminologie normalisée contenue dans la collection des normes internationales de l'ISO et des normes françaises. LEXIS, TEAM, TERMIUM et EURODICAUTOM, ont, à des titres divers, pour but d'assister les traducteurs dans leurs recherches terminologiques effectuées dans une institution ou une société de caractère national ou international.

La spécificité de la clientèle a entraîné des modes de construction et d'opération propres à chacune des banques, des logiciels faits sur mesure et des stocks lexicaux assez différents. Ceci a pour corollaire une fâcheuse incompatibilité de la structure des données.

Suivant la nature de l'unité lexicale traitée, on peut distinguer plusieurs types de banques :

### 1.- Les banques de mots

Le mot est pris au sens d'unité lexicale, simple ou complexe, appartenant au vocabulaire

général;

2.- Les banques de termes ou de terminologies

Le terme désigne le mot d'une langue de spécialité appartenant à un ensemble défini par un champ d'expérience. Sur la suggestion de collègues canadiens, le terme "terminologies" devrait être employé au pluriel, car ce qui est mis en mémoire, ce sont des termes ou des synapsies et non la discipline appelée terminologie;

3.- Les banques de données terminologiques

L'unité terminologique, simple ou complexe, est accompagnée de données sémantiques (définitions, contexte illustratif ou note), de données signalétiques et de données documentaires (la référence ou les sources, le code-matière ou le domaine d'application).

Une *banque de terminologie* n'est pas un simple dictionnaire automatique. Ce dernier est un dictionnaire consultable à distance sur console de visualisation (*télé-dictionnaire*), qui résulte de l'adaptation informatique de dictionnaires existants, l'ordre alphabétique étant devenu numérique. La banque est, au contraire, pour reprendre l'expression de B. Quemada, un *non-dictionnaire*. Nous verrons quels en sont les avantages et les contraintes.

Les banques de terminologie multilingue disposent soit de gros fichiers axés sur l'appariement d'un-terme ou de multitermes en deux ou plusieurs langues, souvent accompagnés de définitions ou de contextes, soit de fichiers contenant des terminologies-objets unilingues assorties de définitions souvent normalisées et accompagnées d'un équivalent dans une autre langue. C'est le cas pour NORMATERM. On distingue :

- 1) Les fichiers unidirectionnels : ils fournissent dans une langue-cible privilégiée, l'équivalent d'un terme d'une ou de plusieurs langues-sources. Dans les deux banques du Canada, c'est l'anglais qui joue le plus souvent le rôle de langue-source.
- 2) Les fichiers multidirectionnels : ils fournissent au départ d'une quelconque langue-source, les équivalents dans une ou plusieurs langues-cibles, toute langue pouvant être, au choix, langue-source ou langue-cible.

Les banques de données terminologiques et lexicales multilingues ont à tenir compte des faits suivants :

- 1) Le lexique constitue une classe ouverte (c'est un ensemble flou) dont le nombre d'unités est illimité.
- 2) Le nombre de signifiants est inférieur au nombre de signifiés : à chaque signifiant correspondent plusieurs signifiés. Le stock lexical a une structure homographique et non une ordonnance polysémique.
- 3) Bien souvent, les lexiques des langues de spécialité ne se correspondent pas de manière

bi-univoque; à un terme correspondent plusieurs termes dans l'autre langue.

- 4) Les équivalences ne se résolvent pas au niveau de l'unité simple, mais de l'unité lexicale complexe.

Exemples :

Lame de scie à ruban à denture unilatérale

Moteur à induction linéaire avec secondaire court mobile

Charrue pour labour à plat à traction animale sans avant-train.

La présente communication se propose de livrer quelques réflexions sur le rôle dévolu à une banque de terminologie mise au service de la traduction. Ce rôle dépend certes d'un bon nombre de facteurs autres que linguistiques : facteurs techniques, informatiques, psychologiques, organisationnels, documentaires.

L'exposé s'impose une double limite : d'une part, décrire une banque qui sert et dessert les traducteurs, d'autre part, n'envisager que le cas de la consultation en direct et en mode dialogué.

## II.- La banque de terminologie au centre du processus de traduction

Il serait simpliste de réduire une banque à une sorte de grange, à laquelle on aurait accès par un terminal et d'où l'on sortirait à la fourche les bottes de foin souhaitées. Une banque de terminologie pour traducteurs vient tout naturellement se greffer sur le processus dynamique de traduction, lequel doit entrer dans le schéma si l'on veut juger de l'efficacité d'une banque de ce genre.

Le processus de traduction comporte trois phases : une phase de compréhension ou d'analyse, une phase de transposition ou de confrontation et une phase de reformulation ou de synthèse. Ces phases sont étroitement liées et interdépendantes.

En amont de la banque, toute évaluation tiendra compte d'un certain nombre de paramètres liés à la nature du texte-source :

- 1) un paramètre lié à la *densité terminologique* du texte-source : le texte peut ne contenir que des termes dénotatifs, monosémiques, tirés d'une nomenclature, et être à ce titre 100% terminologique; il peut n'en contenir qu'un petit nombre;
- 2) un paramètre lié au *choix* des unités terminologiques qui feront l'objet de la question : le traducteur interrogera-t-il la banque, après une première lecture, pour tous les mots dont il n'a pas une connaissance active immédiate, ou seulement, après avoir consulté, sans succès, les moyens traditionnels ?
- 3) un paramètre lié au *découpage* du texte-source en unités terminologiques suivant des critères formels, sémantiques, quantitatifs et taxinomiques;

- 4) un paramètre de neutralisation partielle ou totale du contexte : le traducteur surprend le terminologisme en discours, le décontextualise en quelque sorte et en neutralise l'environnement avant sa recherche.

Lors de son analyse terminologique, le traducteur repère les termes inconnus et découpe le texte-source en unités terminologiques, le plus souvent syntagmatiques.

La statistique révèle que les vocabulaires thématiques usent pour 80% de la combinatoire syntagmatique, qui se manifeste par des collocations en anglais, des compositions en allemand et des synapsies en français.

Exemple :

- D. Druckwasserstoffbeständigkeit : F. résistance à l'hydrogène sous pression
- E. non-dispersive ultra-red gas analysis : F. analyse non dispersive des gaz aux ultra-rouges
- F. charrue pour labour à plat à traction animale sans avant-train :
- D. Gespannkehrpflug ohne Vorderkarre : E. animal drawn reversible plough with front wheel assembly.

Dans la phase de transfert, le traducteur, devenu émetteur-récepteur, entreprend sa consultation en mode dialogué. Il converse avec son interlocuteur digital et confronte avec un texte-cible en devenant les équivalents potentiels proposés hors discours par la banque.

Dans cette phase, plusieurs paramètres interviennent :

- 1) un paramètre d'*acceptabilité* des réponses en fonction du texte-source et du texte-cible;
- 2) un paramètre lié aux possibilités de *combinatoire* syntagmatique.

Relevons ici quelques modèles d'appariement en monosémie et en polysémie.

A. Les cas de monosémie LD

M (LD)  $\equiv$  M (LA) : au niveau de la langue

m (TD)  $\equiv$  m (TA) : au niveau du texte

La correspondance bi-univoque existe au niveau de la langue et le passage peut s'effectuer au niveau du texte sans recours au dénoté : c'est le cas dans la nomenclature chimique, la classification botanique et la taxinomie biologique.

---

M (LD)  $\equiv$  M<sub>1</sub>, M<sub>2</sub>, M<sub>3</sub>, (LA) : niveau de la langue

m (TD)  $\equiv$  m<sub>1</sub> (TA) : niveau du texte

Au niveau de la langue, à un terme monosémique LD correspondent plusieurs synonymes dans LA. Un de ceux-ci est choisi au niveau du texte.

---

$M(LD) \approx M(LA)$  : équivalence possible au niveau de la langue

$m(TD) \equiv m(TA)$  : équivalence acceptable dans le texte

---

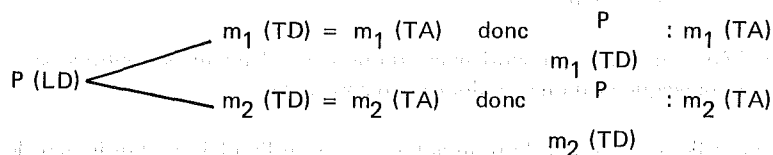
$M(LD) : \phi(LA)$  : pas de correspondance au niveau de la langue

$m(TD) \approx x + y + z(TA)$  : paraphrase au niveau du texte

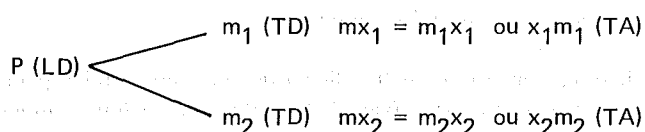
---

### B. Les cas de polysémie LD

La réduction de la polysémie se fait par la mention du domaine d'application ( $x_1$  ou  $x_2$ ).



Réduction de polysémie par collocation ou solidarité lexicale (cotexte syntagmatique) quand le mot apparaît dans le même domaine



Par ailleurs, ce n'est pas uniquement le nombre brut des entités disponibles (stock lexical réel) qui fait la richesse intrinsèque d'une banque. Indépendamment de la pertinence des équivalences, c'est la combinatoire syntagmatique des réponses reçues qu'il faut envisager.

C'est ainsi que des *réponses partielles* peuvent se révéler utiles. Si un multiterme faisant l'objet d'une question ne se trouve pas tel quel, avec tous ses éléments, en mémoire, la machine fournira par programme des éléments de réponse ou réponses partielles à partir desquelles il sera possible, par des interrogations successives, de construire une réponse complète. A la question *aéroglesseur (à) jupes souples*, qui ne se trouve pas telle quelle en mémoire, l'ordinateur cherchera automatiquement sous *aéroglesseur à jupes X*, *X à jupes souples*, *aéroglesseur à X souples*. Dans le cas qui nous occupe, l'interrogateur obtiendra pour l'allemand langue-cible successivement les réponses (1) *Luftkissenfahrzeug mit festen Schürzen* et (2) *Luftkissen mit elastischen Schürzen*, à partir desquelles il pourra construire la réponse totale à sa question, à savoir *Luftkissenfahrzeug mit elastischen Schürzen*.

Ce n'est donc pas le nombre de références qui fait la richesse d'une Banque, c'est la pertinence et la combinatoire des informations et la nature et l'éventail des textes qui sont soumis à l'interrogation.

Des réponses fragmentaires peuvent aussi se révéler utiles et permettre la reconstitution d'une réponse complète qui convient au texte-cible.

On distingue ainsi plusieurs types de réponses : (1) les réponses pertinentes, complètes et partielles; (2) les réponses-propositions; (3) les silences.

Les *réponses pertinentes complètes* sont celles qui correspondent exactement à la question portant sur un uniterme ou sur un multiterme. Elles présupposent une relation bi-univoque pour l'uniterme  $t_s \equiv t_c$ , pour le multiterme  $t_s^1 + t_s^2 = t_c^1 + t_c^2$ , et l'absence d'homographie de  $t_s$ . Celle-ci peut être levée par la mention du domaine d'application.

Les *réponses pertinentes partielles* sont celles qui contiennent un ou deux éléments du multiterme. Elles présupposent une relation univoque et un même domaine d'application.

Les *réponses-propositions* sont celles qui causent du bruit volontaire et où l'on laisse à l'utilisateur le soin de recontextualiser à partir d'une série de réponses possibles. L'équivalence ne peut se résoudre qu'en contexte-cible alors que les données sont fournies et classées en contexte neutralisé.

#### Le silence

La Banque ne contient ni complètement ni partiellement l'information demandée.

Il est trop tôt pour déterminer l'ensemble des paramètres de nature linguistique à mettre en jeu pour évaluer le rôle joué par une banque de terminologie mise au service de la traduction. A tout le moins pourrait-on suggérer quelques grandes directions de recherche.

- 1) La difficulté tient aux rapports complexes qui existent entre le terminologisme plus ou moins neutralisé tel qu'il apparaît dans la banque et le terminologisme qui fonctionne en discours dans le texte-source et qui fonctionnera dans le texte-cible en devenir. A l'exclusion des termes d'une nomenclature (terminologies-objets), l'unité terminologique appartenant à la langue fonctionne nécessairement comme un élément de la phrase. L'analyse terminologique va du terme en discours au terminologisme plus ou moins neutralisé; la synthèse consistera à réinsérer le terme proposé dans le tissu du texte-cible.
- 2) La situation décrite sous (1) conduit à un paradoxe. Plus l'information terminologique mise sur fiche est riche d'indications contextuelles, plus le terme-vedette est enchâssé dans ses axes syntagmatique et paradigmatic et plus son équivalent dans l'autre langue le serre de près, *moins* le terme-vedette se laisse neutraliser et réinjecter dans un texte-cible qui n'est pas purement répétitif.
- 3) En consultation directe, le traducteur a une approche graphémique; pour la machine, le mot est une chaîne de caractères (supports graphiques) et l'interrogation néglige toujours le passage

“aux choses et aux phénomènes”.

- 4) Les paramètres en aval, de transfert et en amont, sont parfois difficiles à discerner d'autant que les pratiques terminologique et traduisante se mêlent souvent en un syncrétisme étroit.