

Les nouveaux lexiques ou une stratégie de création de mini-banques

André Clas

Volume 32, numéro 2, juin 1987

Vers l'an 2000. La terminotique, bilan et perspectives
Objectives: Year 2000 Terminotics. State of the Art, Prospects for the Future

URI : <https://id.erudit.org/iderudit/002250ar>

DOI : <https://doi.org/10.7202/002250ar>

[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

Clas, A. (1987). Les nouveaux lexiques ou une stratégie de création de mini-banques. *Meta*, 32(2), 212–215. <https://doi.org/10.7202/002250ar>

LES NOUVEAUX LEXIQUES OU UNE STRATÉGIE DE CRÉATION DE MINIBANQUES

ANDRÉ CLAS
Université de Montréal

Dans une communication¹ faite au Colloque de l'OLF sur la traduction le 25 octobre 1968, nous déclarions :

Si l'on estime que 50% des connaissances de l'homme de science se perdent au cours de sa carrière à cause de nouvelles découvertes, il devient vital, et même de plus en plus vital, de recourir de plus en plus, et de plus en plus rapidement à la traduction pour faire face à une concurrence de plus en plus difficile, pour obtenir les éléments d'une survie. Ces découvertes entraînent une mise à jour terminologique constante et immédiate.

Et en conclusion, nous demandions que les autorités compétentes

établissent ou fassent établir : — la liste exhaustive des organismes qui s'occupent de recherches terminologiques ; — un comité chargé d'étudier la création d'une **banque de terminologie**. Cette banque de terminologie devrait d'abord coordonner tous les efforts qui se font dans ce domaine. (...) Cette banque aurait donc pour mission de rassembler toutes les recherches terminologiques et de les classer par centres d'intérêt. Au fond, nous suggérons la création d'un dictionnaire automatique, car la banque ne pourrait répondre instantanément à toute demande et posséder une liste de mots constamment à jour que si l'on fournit tous les renseignements à un ordinateur.

À la suite de cette communication, le conférencier fut chargé par les participants du Colloque d'entreprendre les démarches nécessaires pour concrétiser la proposition présentée. L'intéressé convoqua en janvier 1969 une « assemblée » et « quarante sages » répondirent à l'appel. Cette assemblée décida de former un comité d'étude et, en mars de la même année, le rapport soumis par ce comité fut adopté (*META* 14-4, pp. 191-194). Parallèlement, le rapport du comité fut entériné par le Conseil de la faculté des lettres et remis aux autorités de l'Université pour en saisir les gouvernements provincial et fédéral. Une entente tripartite, Secrétariat d'État du Canada, ministères des Affaires culturelles (OLF) et Université de Montréal fut conclue. En 1970, la Banque de terminologie de l'Université fut inaugurée officiellement et les travaux de recherche commencèrent en s'appuyant sur les résolutions du « Colloque de Genève de 1969 » (*META* 15-2, p. 140) qui s'était donné comme thèmes de réflexion : *Recherche terminologique, Documentation terminologique, Formation de terminologues*².

Bien entendu, la Banque de terminologie de l'Université de Montréal (TERMIUM) ne fut pas la première banque du monde : LEXIS (*META* 16-1, 2, pp. 117-124), TEAM (*META* 16-1, 2, pp. 125-131), DICAUTOM (*Journal des traducteurs* 10-3, pp. 78-86) existaient déjà, mais nous croyons qu'elle a ouvert une nouvelle ère de collaboration internationale et a donné une nouvelle impulsion aux travaux à Bruxelles et à Luxembourg (EURODICAUTOM). Mais les choses changent, l'entente tripartite est dénoncée et l'OLF crée, vers 1973, sa propre banque (BTQ). Quelques années plus tard,

TERMIUM est remis aux gouvernements. Le Bureau des traductions exploite TERMIUM et commence à modifier le système pour utiliser, à l'heure actuelle, un produit de 3^e génération. Si nous avons deux « grosses » banques de terminologie au Canada, nous avons aussi quantité de « petites » banques. En effet, la technique a évolué très rapidement et, grâce à la micro-informatique, a rendu accessible à beaucoup de personnes des outils informatiques de plus en plus indispensables dans divers domaines et notamment en terminologie et en traduction.

Les grosses banques suivent l'évolution de la grande informatique et ont leur utilité parce qu'elles contiennent quelques centaines de millions de caractères donnant accès à l'information terminologique ou documentaire, ou même encyclopédique, sur un grand nombre de domaines spécialisés et pour un grand nombre de spécialistes des secteurs de toute la connaissance humaine. Elles constituent en quelque sorte un « trésor » de terminologies.

Les mini-banques ont des objectifs beaucoup plus modestes. Elles ne visent plus la collectivité, mais l'individu. Le stade industriel redevient artisanal ! Les mini-banques suivent l'évolution de la micro-informatique et sont donc conçues pour fonctionner sur un micro-ordinateur d'une capacité mémorielle interne de 256 ou 512 kilooctets et d'une mémoire auxiliaire d'environ un mégaoctet. Bien entendu, les mini-banques ne peuvent pas concurrencer les grosses banques, elles ne le doivent pas non plus, leurs rôles étant autres, car elles répertorient la terminologie dans un seul domaine spécialisé. En fait, les mini-banques sont appelées à remplacer les dictionnaires, glossaires, lexiques et vocabulaires spécialisés du traducteur par des disquettes contenant la même information générale, mais multipliant les renseignements par un accès immédiat, diversifié et matriciel. En outre, ces mini-banques peuvent être mises à jour sans délai. La bibliothèque du traducteur-terminologue devient sa « disquethèque ». Certains préfèrent les livres et se refusent au changement à cause de leur « amour » pour l'ancien « serviteur », mais peut-on arrêter le progrès, surtout s'il est bénéfique à tous égards ?

Bien entendu, divers modèles de mini-banques sont possibles et le modèle à retenir est celui qui correspond le mieux aux besoins des utilisateurs. Ces besoins sont essentiellement la découverte de l'équivalent recommandé dans l'autre langue du terme recherché ou, à défaut de réponse valable, la présentation d'une matrice utilisable parce qu'elle livre des données proches sémantiquement ou structurellement. En plus de fournir le simple équivalent, une banque devra donner toutes les possibilités d'accès à des synonymes, à des abréviations, aux lexies simples des lexies composées et des lexies complexes. Ainsi, **trempe à l'air** (E *water hardening*, D *Wasserhärten*), **trempe à l'huile** (E *oil hardening*, D *Ölhärten*), **trempe au plomb** (E *lead hardening*, D *Bleibadhärten*), **trempe en bain de sel** (E *salt bath hardening*, D *Salzbadhärten*), **trempe par aspersion** (E *spray hardening*, D *Spruhhärten*), **trempe en coquille** (E *chill hardening*, D *Quettenhärten*), **trempe au chalumeau** (E *flame hardening*, D *Flammhärten*), **trempe par induction** (E *induction hardening*, D *Induktionshärten*), **trempe à cœur** (E *full hardening*, *core hardening*, *penetration hardening*, D *Durchhärtung*), **trempe superficielle** (E *surface hardening*, D *Oberflächenhärtung*, *Einhärtung*), **trempe au passage**, **trempe de proche en proche** (E *continuous hardening*, *progressive hardening*, D *Durchlaufhärten*, *Umlaufhärten*, *Vorschubhärten*). Tous ces termes devront aussi renvoyer au premier, mais pour éviter une avalanche, la micro-banque, en plus de donner l'équivalent du terme demandé, donnera des renseignements du type suivant : « il y a d'autres formations avec la lexie simple ». Pour les lexies complexes du type **radiophare d'alignement de descente** (E *glide path beacon*, *glide path localizer*, D *Landerichtungsbake*) ou **alimentation des véhicules électriques par invertisseur** (E *power supply to electric vehicles by the inverter system*, D *Speisung der Elektrofahrzeuge durch das Invertor-System*), il faut une décompo-

sition en lexies simples si l'on veut des possibilités d'accès à des renseignements sémantiques et matriciels. Ainsi, en plus des termes complets, il faudra retenir les lexies simples suivantes :

- ◆ pour **radiophare d'alignement de descente** :
radiophare — alignement — descente — radiophare d'alignement — alignement de descente ;
- ◆ pour **alimentation des véhicules électriques par invertisseur** :
alimentation — véhicule — invertisseur — alimentation de véhicules — véhicules électriques ;
- ◆ pour **glide path beacon**,
le synonyme *glide path localizer* et les lexies simples : *glide — path — beacon — glide path — localizer* ;
- ◆ pour **power supply to electric vehicles by the inverter system**, les lexies simples :
power — supply — vehicle — inverter — system — power supply — electric vehicle — inverter system ;
- ◆ pour l'allemand, il faudra décomposer le terme et lemmatiser les lexies obtenues, ainsi on aurait pour *Landerichtungsbake — Landen — Richtung — Bake — Landerichtung*, et pour *Speisung der Elektrofahrzeuge durch das Inverter-System*, on aurait *Speisung — Fahrzeuge — Inverter — System — Elektrofahrzeuge — Speisung der Elektrofahrzeuge — Inverter System*.

Toutes ces lexies constituent des clés d'accès aux termes. Bien entendu pour garantir la cohérence de contenu et donc la systématique des données, il faut déterminer une procédure de décomposition. La méthode la plus simple, du moins pour le français et l'anglais, consiste à retenir les noms, les noms composés, les adjectifs, les verbes, les adverbes, les codes. Comme on le voit, les termes composés ou complexes génèrent toute une série de lexies qu'il s'agit d'intégrer à la mini-banque et il faudra trouver la meilleure façon pour éviter un encombrement et un débordement rapide des mémoires de l'ordinateur. Pour bien structurer la banque, des calculs statistiques sont fondamentaux. L'analyse de lexiques permet une première estimation qui peut sembler banale, mais qui est néanmoins indispensable, car elle conditionne en partie l'espace mémoire disponible et détermine ainsi les possibilités de présentation d'autres données. On peut ainsi estimer qu'un lexique spécialisé a en moyenne 3 500 termes et la répartition statistique par catégories grammaticales révèle que 94% sont des substantifs, 3,4% des verbes ou locutions verbales avec des verbes opérateurs généraux, 1,7% sont des adjectifs et 0,6% sont des codes. Bien entendu, on sait depuis toujours que la dénomination est surtout faite à l'aide de substantifs, mais ces substantifs sont de divers types : environ 8% des termes sont des monominaux, 34% des termes sont du type N + adj, 35% des termes sont des binominaux du type N + N ou N prép N, qui peuvent encore avoir une expansion adjectivale, 7% des termes sont des trinominaux du type N + N + N ou N prép N prép N et environ 2% des termes sont des « composés lourds » (formations périphrastiques).

Ces statistiques sont d'un grand intérêt pour déterminer l'espace mémoire nécessaire et par des comparaisons fixer une stratégie de stockage. Une autre donnée nécessaire est le calcul de la longueur moyenne d'une lexie, c'est-à-dire le nombre de mots et de caractères par terme, et la proportion de mots différents qui les constituent. Une étude (Baudot 1986b : 6-12) stipule qu'on peut estimer le nombre de mots différents pour un lexique de 2 000 termes à 37% du nombre total de mots, que cette « proportion de mots différents tombe à 18% pour un lexique de 16 000 termes, et à 16% pour un lexique de 20 000 termes ». La longueur moyenne d'un mot est de l'ordre de 7,5 caractères (Baudot 1986a : 25).

Ces analyses et ces statistiques permettent de déterminer un mode de stockage optimal des informations. En plus du stockage des fiches terminologiques, c'est-à-dire du terme, de ses synonymes et de divers renseignements, il faut stocker les tables et les index. Le stockage des fiches terminologiques se fait en mémoire auxiliaire, c'est-à-dire sur disquettes, puisqu'il n'y a nécessité d'accès qu'après le traitement d'une requête de l'utilisateur. Par contre, la table des termes, c'est-à-dire la liste de tous les termes dans la banque, tout comme l'index, c'est-à-dire la liste des clés d'accès aux fiches terminologiques seront stockés en mémoire interne, permettant ainsi un accès rapide à l'information cherchée.

Pour « gagner » de l'espace mémoire, il est très rentable de stocker une table des termes et d'y transcoder toutes les lexies par des numéros. On constate que les tables — table de lexies différentes et table de termes transcodées —, occupent non seulement moins d'espace mémoire, mais cette réduction est encore plus avantageuse lorsque le nombre de termes augmente. Ainsi, on peut estimer que la réduction atteint 37% pour 18 000 termes et passe à 47% pour 20 000 termes (Baudot 1986b : 6-11).

La constitution d'un index et sa conservation en mémoire interne sont nécessaires pour donner un accès rapide aux fiches terminologiques. Même si le nombre de lexies simples est important, on constate qu'elles sont souvent répétées et donc ne sont à transcoder qu'une seule fois. Il suffit d'étiqueter les entrées pour orienter l'accès à l'information. Pour réduire encore l'encombrement de la mémoire, il est intéressant de changer l'index linéaire en index de structure arborescente et de ne pas répéter pour chaque entrée toutes les lettres identiques à celles de l'entrée précédente. Cette façon de faire est très économique pour des composés lourds qui, dans un même domaine, commencent souvent par les mêmes lexies simples.

La démarche que nous proposons montre toute l'utilité de la constitution d'une mini-banque de terminologie fonctionnant sur micro-ordinateur individuel pour le traducteur. La rapidité, la souplesse de la consultation, la mise à jour facile sont des atouts déterminants pour assurer le succès de ces lexiques informatisés.

Notes

1. Nous reprenons partiellement le texte de cette communication parce que ce texte n'a jamais été publié, parce qu'il est le point de départ des « banques » canadiennes et parce qu'il nous semble encore d'actualité.
2. Le *Petit Robert 1* date ce mot du mil. XX^e ; au Québec. Est-ce que cette attestation est une datation plus précise et une origine différente ?

BIBLIOGRAPHIE

- BAUDOT, Jean (1986a) : « Structure informatique interne d'une mini-banque de terminologie », *la Banque des mots*, n° 31, Paris, CILF, pp. 19-27.
- BAUDOT, Jean (1986b) : *Rapport de recherche* : « BATEM : Développement et mise au point d'un système de petite banque de terminologie spécialisée fonctionnant sur micro-ordinateur », GRESLET, Université de Montréal, pp. 6-1 à 6-16.
- CLAS, André et Jean BAUDOT (1986) : « BATEM. Une banque de terminologie sur micro-ordinateur », *Infoterm Series 8 : Deuxième symposium*, München, K.G. Saur, pp. 376-389.