

Les banques de terminologie de l'avenir

Jean Baudot

Volume 31, numéro 2, juin 1986

URI : <https://id.erudit.org/iderudit/002004ar>

DOI : <https://doi.org/10.7202/002004ar>

[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

Baudot, J. (1986). Les banques de terminologie de l'avenir. *Meta*, 31(2), 153–158.
<https://doi.org/10.7202/002004ar>

LES BANQUES DE TERMINOLOGIE DE L'AVENIR*

JEAN BAUDOT
UNIVERSITÉ DE MONTRÉAL

1. INTRODUCTION

Il est peut-être utile de rappeler les principales raisons qui ont motivé l'utilisation de l'informatique dans la gestion des données terminologiques.

On sait qu'il est rarement économique d'utiliser l'informatique dans une activité qui peut s'effectuer par des moyens classiques. Par exemple, informatiser un dictionnaire usuel, tout en maintenant l'usage habituel qu'on en fait, n'aurait pas d'intérêt et ne représenterait en rien une économie. Ce n'est que lorsque l'informatique nous permet de faire des choses impossibles sans elle que son utilisation devient justifiable.

Dans les cas de fichiers de données terminologiques, il s'agit de traiter des masses considérables de données, qui atteignent facilement les centaines de millions de caractères, pour répondre à plusieurs exigences :

- ◆ ces données doivent avoir une structure multidimensionnelle : langues, domaines, sources, organismes, etc. ;
- ◆ on doit pouvoir avoir accès à ces données selon différents modes de consultation ;
- ◆ enfin, on doit pouvoir interroger ou mettre à jour le fichier de plusieurs points, et cela simultanément.

Il n'y a pas de doute que seul un fichier informatisé puisse répondre à ces exigences.

2. HISTORIQUE

Depuis les premiers travaux de mise sur pied de banques de terminologie informatisées, il y a maintenant déjà presque 20 ans de cela, il n'y a pas vraiment eu de changement majeur ni dans la constitution ni dans le mode d'utilisation des banques de termes. Tout au plus, peut-on faire état d'une évolution qui s'est effectuée très rapidement au début des années 70 grâce aux développements rapides de l'informatique conversationnelle : celle du passage de l'interrogation en différé à l'interrogation en direct. Grâce à cela, l'utilisateur ne devait plus attendre 24 heures et pouvait obtenir ses renseignements terminologiques immédiatement, ou ... presque.

Certes, on a ajouté quelques à-côtés aux systèmes d'interrogation informatisés pour en élargir la rentabilité, comme le couplage à un système de traitement de textes qui constitue une première étape vers la transformation du terminal d'interrogation en poste de travail pour l'utilisateur, ou encore la production automatique de lexiques sectoriels à partir d'un balayage partiel ou systématique du contenu bancaire.

Il n'y a pas de doute que, si l'on tente de faire aujourd'hui le point sur l'utilisation des banques de terminologie, le bilan est positif et nous savons tous combien sont précieux ces réservoirs de renseignements terminologiques. Pourtant nous savons aussi qu'il y a des limites et des faiblesses dans les systèmes existants et que ceux-ci ne répondent pas toujours à toutes nos attentes.

Tout d'abord, ces systèmes coûtent cher et les banquiers le savent. Il faut dire aussi que la plupart des banquiers se sont laissés convaincre par leurs conseillers informaticiens d'utiliser, pour informatiser leurs banques, des progiciels de gestion de bases de données préfabriqués. Il est vrai que, d'une part, les spécialistes en terminologie ne sont pas toujours en mesure d'évaluer les avantages et les inconvénients des différentes ressources informatiques matérielles et logicielles et que, d'autre part, les informaticiens ne connaissent pas toujours toutes les caractéristiques des données terminologiques. Cette situation est courante dans l'informatisation des grandes bases de données et ne se limite pas aux banques de terminologie. Ce que j'appelle les systèmes de gestion de bases de données préfabriqués sont des systèmes qui tentent de répondre à des besoins très variés, parfois très différents et qui, pour cette raison, sont souvent des systèmes très encombrants à cause de la généralité et de l'adaptabilité qu'ils doivent obligatoirement conserver. Je persiste à croire que ce n'est pas nécessairement la meilleure solution lorsqu'une banque de données particulières doit être utilisée très fréquemment par un grand nombre de personnes. Je suis bien conscient cependant que l'alternative, celle qui consisterait à développer un système *ad hoc*, n'est pas toujours envisageable pour des raisons d'ordre pratique : délai de mise en service, responsabilité de soutien et d'entretien du système, disponibilité d'une équipe compétente, etc. Il n'en reste pas moins que le système préfabriqué est un compromis qui, s'il se justifie selon certains critères, est souvent plus rentable pour les fabricants de matériel que pour les utilisateurs.

Une autre insuffisance des banques de terminologie réside dans leur accessibilité. Les banques de terminologie ne sont pas aujourd'hui véritablement accessibles à tous ceux qui pourraient en bénéficier, eussent-ils un terminal à leur disposition. Cette faiblesse est beaucoup plus d'ordre technique et son remède dépendra de développements technologiques ainsi que d'une réduction des coûts de production d'appareils et des frais de télécommunications. Je ne crois pas que le terminal domestique polyvalent puisse déjà donner accès à des informations terminologiques à très peu de frais. Tout au plus peut-on souhaiter que les publications automatisables de recueils terminologiques soient intensifiées tant sur la base de l'étendue sectorielle que sur celle de la périodicité des mises à jour.

3. DEUX VOIES DE DÉVELOPPEMENT

Mais alors que peut-on attendre ou envisager dans les prochaines années ? Les progrès informatiques vont-ils modifier la conception et le fonctionnement des banques de terminologie ? Dans quelles directions les banques de terminologie vont-elles évoluer ?

Je crois, pour ma part, que deux voies de développement se dessinent, deux voies qui demeureront distinctes pendant quelque 10 ou 20 ans, mais qui convergeront inévitablement.

La première voie aura pour objectif de mieux renseigner, mieux servir l'utilisateur humain : le traducteur, le rédacteur, tous ceux qui ont des préoccupations d'expression langagière correcte et appropriée. On peut donc entrevoir dans cette voie des banques, d'une part plus compactes et portables, d'autre part plus complètes, soit par le multilinguisme, les illustrations ou même le mouvement.

L'autre voie de développement vise la clientèle des machines informatiques elles-mêmes. Nombreux sont les systèmes informatiques qui auront besoin d'informations terminologiques dans la base de connaissances informatisées sur laquelle ils comptent s'appuyer. Ni les banques de terminologie actuelles ni les dictionnaires présentement informatisés ne peuvent encore répondre à ces besoins. On devrait donc, dans cette deuxième voie, assister à l'apparition de véritables dictionnaires électroniques qui cons-

titueront les bases de données terminologiques des systèmes (artificiellement) intelligents.

4. LES PROCHAINES BANQUES DE TERMINOLOGIE

Parmi les prochaines banques de terminologie, celles qui sont le plus susceptibles de se répandre rapidement sont les minibanques ou banquetteres. Des développements dans cette direction sont en cours présentement en Europe, aux États-Unis et au Canada. Comme son nom l'indique, une minibanque est une banque réduite destinée à être utilisée sur un ordinateur personnel. La réduction de la banque est nécessaire pour ne pas dépasser les possibilités de stockage des microordinateurs actuels. En effet, si une grande banque de terminologie peut facilement atteindre plusieurs centaines de millions de caractères, une minibanque raisonnable et portable ne peut aujourd'hui guère dépasser le million de caractères. La réduction ne peut se faire qu'à deux niveaux : d'une part en limitant une minibanque à un seul domaine ou sous-domaine de spécialité, d'autre part en réduisant la taille des fiches terminologiques. Une minibanque monosectorielle peut ainsi facilement contenir quelque dix mille fiches bilingues ou multilingues contenant chacune plusieurs renseignements terminologiques. Ceci dépasse la taille de bien des dictionnaires de spécialité.

Certes, une minibanque ne contiendra vraisemblablement jamais tous les renseignements auxquels on peut avoir accès dans une grande banque de terminologie, mais la vocation des minibanques est bien différente de celle des grandes banques. Les minibanques constituent en quelque sorte des dictionnaires ou lexiques spécialisés informatisés avec le double avantage de permettre l'inclusion de certains renseignements terminologiques et d'offrir des moyens d'accès à l'information beaucoup plus raffinés que ne le permettent les ouvrages imprimés. Physiquement, une minibanque se présente sous la forme d'une ou deux disquettes et on peut déjà prévoir que les disquettes remplaceront progressivement les ouvrages imprimés sur les étagères des utilisateurs.

Ainsi, le traducteur qui entreprend la traduction d'un texte dans un certain domaine, n'aura qu'à charger la minibanque pour ce domaine sur son ordinateur personnel. Dans les cas où la minibanque s'avère insuffisante, l'utilisateur pourra toujours interroger une grande banque en se servant de son ordinateur comme terminal d'interrogation. Ce mode de travail présente le double avantage d'offrir un accès plus rapide à une terminologie sectorielle et de réduire les coûts de télécommunication.

Il existe une autre façon d'envisager la mise en disponibilité d'une minibanque sur un ordinateur personnel : le téléchargement sélectif à partir d'une grande banque de terminologie. Dans ce scénario, le traducteur qui entreprend la traduction d'un texte dans un certain domaine utilise son ordinateur comme terminal pour se brancher sur une grande banque. Une fois connecté, il commande alors le chargement d'un sous-ensemble approprié de la grande banque dans la mémoire de son ordinateur. Ce sous-ensemble est défini à la fois par une limite sectorielle et par une description de sous-fiche. Une fois chargée, la minibanque peut alors être consultée sur place. L'utilisateur peut ainsi charger les minibanques dont il a besoin au cours de son travail.

Il y a toutefois une ombre au tableau de ce deuxième scénario. En effet, le téléchargement d'une minibanque représente une grande quantité d'informations à transmettre et rendrait peu pratique ce mode d'opération avec les débits de transmission de données disponibles sur les réseaux de télécommunication actuels. Il faudra attendre quelques années avant de voir se répandre cet autre type de minibanque. Peut-être même n'aura-t-il même pas le temps de prendre de l'importance car un autre progrès technologique risque fort de devenir beaucoup plus populaire : il s'agit des disques optiques.

Les disques optiques constituent un important progrès dans le stockage d'informations en mémoire auxiliaire des systèmes informatiques. Ils offrent en effet quatre avantages majeurs sur les disques magnétiques rigides habituels :

- ♦ premièrement, leur capacité de stockage est au moins dix fois plus élevée pour des disques de tailles similaires ;
- ♦ deuxièmement, un disque optique avec l'information qu'il contient peut être reproduit en série à peu de frais, ce qui n'est pas le cas pour un disque magnétique rigide ;
- ♦ troisièmement, le disque optique est interchangeable, comme la disquette d'un microordinateur ;
- ♦ enfin, le disque optique assure une intégrité exceptionnelle des données ; on parle d'une permanence des informations enregistrées de l'ordre de 50 ans.

Ils ont par contre deux désavantages par rapport aux disques magnétiques rigides :

- ♦ d'une part, le temps d'accès à l'information est environ dix fois plus long ; il est du même ordre de grandeur que celui du temps d'accès à une disquette ;
- ♦ d'autre part, l'information enregistrée ne peut être modifiée. Il est vrai qu'on prédit des disques modifiables d'ici trois à cinq ans, mais ce n'est pas encore le cas des systèmes commercialisés aujourd'hui.

Ces caractéristiques sont d'un intérêt considérable pour les banques de terminologie. En effet, un disque optique peut contenir plusieurs centaines de millions de caractères et peut être branché à un microordinateur. La capacité d'un disque optique est donc du même ordre de grandeur que celui de la taille des grandes banques. Si l'on y enregistre des informations terminologiques il ne s'agirait plus de minibanques, mais bien de copies quasi intégrales des grandes banques.

Pourquoi donc ne pas enregistrer une grande banque sur un disque optique, le reproduire à quelques milliers d'exemplaires et vendre ceux-ci à un prix raisonnable ? La raison en est que l'information est une chose et l'accès à cette information en est une autre. Il ne suffit pas d'avoir une encyclopédie sur un disque optique, encore faut-il que cette information soit structurée de façon adéquate, que l'on dispose d'un logiciel d'interrogation approprié et qui puisse tourner sur un ordinateur personnel, et qu'enfin le lecteur de disque optique et son interface soient d'un prix abordable pour le détenteur d'un microordinateur. Ces obstacles seront sûrement surmontés dans un avenir assez rapproché et je suis étonné qu'il n'y ait pas encore de grand dictionnaire de cette forme sur le marché. Il faut mentionner cependant que certaines encyclopédies américaines ont choisi cette option (par ex. : *Grolier Academic American Encyclopedia*).

Il est certain que, tôt ou tard, des terminologies deviendront disponibles sur disque optique, car ce médium présente un autre avantage considérable : celui de pouvoir stocker à la fois des images et des informations textuelles. On peut donc déjà prévoir trois générations de banques de terminologie sur disques optiques :

- ♦ une première génération constituée uniquement de matériel textuel ;
- ♦ une deuxième, illustrée d'images et de schémas ;
- ♦ et une troisième contenant également des séquences d'animation permettant d'afficher sur écran des illustrations dynamiques.

De là à l'encyclopédie illustrée, multilingue, interrogeable de multiples façons, il n'y a qu'un pas qui sera sûrement franchi avant l'an 2000.

5. LE DICTIONNAIRE ÉLECTRONIQUE

Mais revenons au dictionnaire électronique. Tous les développements dont nous venons de parler pour les banques de terminologie satisferont de mieux en mieux les besoins d'informations terminologiques des personnes. Par contre, les systèmes informati-

ques dont on annonce les progrès en intelligence artificielle auront, eux aussi, de plus en plus besoin d'informations terminologiques. Or, ces systèmes ne trouveront pas dans les banques de terminologie, telles que nous les concevons aujourd'hui, les renseignements dont ils auront besoin. Ils ne les trouveront pas non plus dans les dictionnaires présentement informatisés. Ceux-ci ne sont d'ailleurs le plus souvent qu'une copie du matériel textuel sur support magnétique pour en faciliter la mise à jour et la photocomposition.

Quelles sont donc alors les informations qui devront figurer au dictionnaire électronique ? Je ne suis pas en mesure de répondre de façon satisfaisante à cette question car les besoins ne sont pas encore suffisamment connus. Tout au plus peut-on esquisser quelques aspects évidents de la question.

Les dictionnaires électroniques sont destinés à des utilisations informatiques. Ils auront la forme de tables auxquelles des programmes divers auront accès.

Considérons un exemple très simple : celui du vérificateur d'orthographe. Un tel système, s'il se veut quelque peu efficace, aura besoin d'une foule de renseignements dont la plupart ne se trouvent présentement ni dans les dictionnaires informatisés, ni dans les banques de terminologie. Ces renseignements devront décrire les comportements morphologiques, syntaxiques, et même sémantiques, des entrées figurant dans les répertoires. Et que dire des mots ou termes corrects qui ne figurent pas dans ces répertoires ? De nombreuses dérivations sont si régulières que les dictionnaires ne les répertorient pas toutes et laissent au lecteur le soin de reconnaître les règles de dérivation. Par exemple, le *Larousse* et le *Robert* listent « accrocher », « accrochage » et « accrocheur », mais ne listent pas « accrochable » et il en est ainsi pour de nombreuses dérivations acceptables avec divers affixes. Le dictionnaire devra donc aussi contenir des renseignements grammaticaux sur les règles de dérivation.

Un autre exemple illustrera combien se compliquent rapidement un mécanisme pourtant simple : celui du pluriel des noms. Aucune banque de terminologie, à ma connaissance, ne fournit de renseignements complets sur le pluriel des groupes nominaux en français. Il faut dire que le problème est loin d'être simple et que ni les grammairiens ni les spécialistes ne sont entièrement d'accord sur le sujet. Voici, pour illustrer ce phénomène, quelques termes binominaux anglais du domaine de l'informatique avec leurs équivalents en français. Si la mise au pluriel est élémentaire en anglais, elle est loin de l'être en français :

<i>type declarations</i>	déclarations de type(s)
<i>program units</i>	unités de programme(s)
<i>memory units</i>	unités de mémoire(s)
<i>base types</i>	types de base
<i>procedure calls</i>	appels de procédure(s)
<i>loop parameters</i>	paramètres de boucle(s)
<i>object modules</i>	modules objets

Dans certains cas, le deuxième substantif prend toujours le pluriel, dans d'autres jamais, dans d'autres encore, le pluriel semble facultatif, dans d'autres enfin, cela dépend du contexte et ce sont alors des critères sémantiques qui doivent intervenir. Or, on sait que beaucoup de termes techniques ont une structure plus complexe que ce simple patron binominal. Comment ces renseignements, pourtant élémentaires, seront-ils codés dans les dictionnaires électroniques de l'avenir ? Ce sont là des questions auxquelles les bâtisseurs des futures banques de terminologie devront apporter des solutions. On peut imaginer déjà que chaque terme puisse être décrit par un vecteur linéaire caractérisant chacun de ses composants. Dès lors, des transformations comme la mise au pluriel d'un groupe nominal, la mise au féminin d'un groupe adjectival, la conjugaison d'un groupe verbal, pourraient être effectuées par des opérateurs appropriés agissant sur ces

vecteurs. Mais c'est là de la prospective, car bien d'autres renseignements devront être codés dans les dictionnaires électroniques.

6. CONCLUSION

Il est clair en tout cas que ces deux types de répertoire — dictionnaires informatisés et banques de terminologie d'une part, dictionnaires électroniques d'autre part — n'ont en commun que les vedettes de leurs rubriques respectives. Dans l'avenir immédiat on peut s'attendre à ce que ces différences s'accroissent ; par contre, la systématisation qu'on apportera dans la méthodologie de constitution de ces deux types de répertoires devrait, à moyen ou à long terme, mener à un rapprochement progressif et à une convergence vers un point commun : l'encyclopédie électronique.

Note

* Colloque OLF-STQ : *l'Ère nouvelle de la terminologie*, Montréal, 27-29 novembre 1985.