

La mise sur DOC d'une banque de données linguistiques : l'exemple de TERMIUM

Sylvie Dupont

Volume 34, numéro 3, septembre 1989

1. Actes du Colloque Les terminologies spécialisées : Approches
quantitative et logico-sémantique et 2. Actes du Colloque
Terminologie et Industries de la langue

URI : <https://id.erudit.org/iderudit/002678ar>

DOI : <https://doi.org/10.7202/002678ar>

[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

Dupont, S. (1989). La mise sur DOC d'une banque de données linguistiques :
l'exemple de TERMIUM. *Meta*, 34(3), 594–596. <https://doi.org/10.7202/002678ar>

LA MISE SUR DOC D'UNE BANQUE DE DONNÉES LINGUISTIQUES : L'EXEMPLE DE TERMIUM

SYLVIE DUPONT
*Secrétariat d'État
Ottawa, Canada*

Dans un pays officiellement bilingue et multiculturel comme le Canada, il ne faut pas s'étonner de l'importance du rôle que joue la traduction dans le développement du pays. Ainsi, le ministère du Secrétariat d'État abrite un centre de spécialisation linguistique de haut calibre qui englobe les domaines variés de la traduction, de l'interprétation et de la terminologie. Essentielles à l'appui du bilinguisme institutionnel, ces activités permettent à l'État d'atteindre les objectifs qu'il s'est fixé en matière de langues officielles et de communication en langues étrangères.

À cette fin, le Secrétariat d'État exploite, depuis 1974, une banque de données linguistiques, TERMIUM, déjà bien connue dans les milieux de la traduction et de la terminologie, puisqu'elle compte près de 3 000 utilisateurs tant en Amérique du Nord qu'en Europe, et même en Nouvelle Calédonie !

Tout comme la langue elle-même et la terminologie qu'elle emploie, cette banque de données linguistiques est en constante évolution. À cet égard, le Secrétariat d'État étudie actuellement une nouvelle version de TERMIUM sur DOC, ou disque optique compact.

Ce disque optique ressemble au disque compact audio que l'on connaît déjà. Sur ce petit disque de 12 centimètres, l'information, gravée par faisceau laser, y est stockée sous forme numérique, c'est-à-dire sous forme de 0 et de 1. Un autre faisceau laser permet ensuite de lire le contenu du disque grâce à un lecteur branché à un micro-ordinateur avec une interface adéquate.

Compte tenu de sa petite taille, la capacité de mémoire est inouïe : 550 mégaoctets ou millions de caractères, soit l'équivalent de 1 400 disquettes standard, ou encore l'équivalent de 1 000 livres de poche !

Au cours des dernières années, l'usage du DOC s'est rapidement étendu en Amérique du Nord, où des documents volumineux, tels encyclopédies et dictionnaires ont été versés sur ce type de support. Cette technologie a d'ailleurs le mérite d'être uniformisée à l'échelle internationale.

Parmi les avantages de ce disque compact, notons qu'il libère l'utilisateur qui peut ainsi consulter la banque de données aux heures qui lui conviennent sans être tributaire de l'horaire d'un ordinateur central qui doit être fermé périodiquement pour la mise à jour, la maintenance et les réparations.

Les nombreux fuseaux horaires, dont six au Canada, sont d'ailleurs l'un des facteurs qui militent en faveur de la consultation de TERMIUM par d'autres moyens que l'interrogation en direct.

La facilité d'accès est également un atout. Aucune commande d'entrée en communication n'est nécessaire : inutile de mémoriser de code d'utilisateur ni de mot de passe. De

plus, si le micro-ordinateur est programmé en conséquence, il est possible de lancer automatiquement le logiciel de consultation dès qu'on allume le micro.

Dans le cas d'un micro-ordinateur dédié à la consultation de TERMIUM, on peut le mettre en marche le matin et le laisser allumé pendant des heures d'affilée. Cette autonomie du poste de travail est aussi très avantageuse côté coûts : puisqu'il n'est plus nécessaire de passer par un réseau de télécommunications, il n'y a aucun coût de branchement. Seule l'utilisation du micro-ordinateur à d'autres fins limite l'utilisateur dans sa consultation de TERMIUM. En d'autres termes, l'utilisateur est entièrement autonome : la consultation est illimitée quant au nombre de termes interrogés et au nombre d'heures d'utilisation.

Mais revenons un peu en arrière...

C'est vers la fin de 1985 qu'une compagnie torontoise proposait au gouvernement canadien de verser TERMIUM sur DOC en utilisant son logiciel de consultation. Après de nombreux tests préliminaires effectués sur un échantillonnage de quelque 500 000 fiches, certaines modifications ont été apportées, et une première version de TERMIUM était saisie sur disque optique compact.

Au cours de l'automne 1987, un projet-pilote de trois mois a ainsi été initié pour évaluer la réaction des utilisateurs de TERMIUM à cette nouvelle technologie, ainsi que pour identifier les besoins spécifiques d'une telle application. Au cours de cet essai, la version expérimentale fut installée dans dix sections de traduction du gouvernement fédéral ainsi que dans dix entreprises privées. Ces traducteurs ont ensuite été encouragés à n'interroger TERMIUM qu'à partir de cette version plutôt qu'en mode direct.

De plus, ces participants ont répondu à un questionnaire d'appréciation, fourni des données statistiques sur leurs séances de travail, et soumis périodiquement des listes de termes interrogés, pour fins de comparaison à la version en direct, en vue d'évaluer la fréquence nécessaire de mise à jour du disque optique.

Ainsi, en interrogeant, huit mois après la gravure du disque, les 1 148 termes ainsi relevés, on a retrouvé une réponse différente en banque dans 116 cas. Il s'agissait de fiches qui avaient été créées, modifiées ou annulées depuis mars 1987.

On put ainsi établir que le contenu de la banque centrale varie à un taux de 1,26% par mois. Ainsi, en remettant aux abonnés un nouveau disque tous les quatre mois, soit trois fois par année, seul 5% du contenu du disque pouvait différer de la version du système central entre chaque nouvelle copie.

Bref, l'expérience s'est révélée un tel succès, qu'en octobre dernier, le Secrétariat d'État offrait à toutes les personnes et à toutes les entreprises déjà intéressées par TERMIUM, la possibilité de participer au deuxième volet du projet.

Le contenu de TERMIUM a donc été mis sur DOC en décembre et 250 nouvelles copies seront distribuées sous peu. Ce projet doit durer quelque dix-huit mois, après lesquels une nouvelle version de TERMIUM sur DOC sera disponible à plus grande échelle.

En fait, TERMIUM sur disque optique compact, c'est 885 000 fiches en anglais et en français, soit 3 millions de termes, répartis en trois fonds :

- un fonds de données terminologiques ;
- un fonds de difficultés de traduction et de problèmes de langue ;
- et un fonds d'appellations officielles, qui regroupe les noms d'associations, titres de lois, marques déposées, noms géographiques, etc.

Le fonds de données multilingues qui regroupe des fiches en espagnol et en allemand n'a pas été retenu pour cette version.

Tous les domaines de la connaissance sont traités, mais à des degrés différents en raison du milieu de travail où a évolué TERMIUM. Ainsi, on relève quelque 100 000

fiches dans les domaines de l'industrie manufacturière, 90 000 fiches dans le domaine des transports, 48 000 fiches dans les sciences de l'environnement, 33 000 fiches en informatique, 78 000 fiches en économie et ainsi de suite.

Chacune de ces fiches traite d'une notion précise et, en plus du domaine d'application, comporte une bannière qui indique à quel fonds (terminologie, traduction, ou appellations) et à quel fichier, A, B ou C, selon le degré de certitude des équivalences, appartient la fiche. Le contenu linguistique est ensuite présenté en format parallèle, comme c'est actuellement le cas dans la version en direct. On y retrouve les termes, suivis de leurs synonymes, variantes et abréviations (s'il y a lieu), tant en anglais qu'en français.

La fiche comporte aussi bien d'autres renseignements : chaque terme est pondéré (c'est-à-dire qu'il est jugé CORRECT ou À ÉVITER), et accompagné des justifications d'usage, tel des définitions, des contextes explicatifs et des observations, techniques ou linguistiques.

On y retrouve également la liste de tous les ouvrages consultés et retenus au cours de la recherche, la date de création de la fiche et, le cas échéant, la date de la plus récente mise à jour avant saisie sur DOC.

Cette version TERMIUM comporte également quatre index. Un index est une liste alphabétique de mots ou de termes, en français ou en anglais, tels qu'ils apparaissent sur la fiche : la sélection d'un terme appelle la fiche correspondante.

Le premier index comporte tous les termes ou appellations officielles apparaissant sur la partie française de la fiche, y compris synonymes, variantes et abréviations. Ainsi, le terme *linguistique* apparaît entre les autres entrées *linguiste* et *linguiste appliquée*.

Le deuxième index comporte tous les termes ou appellations officielles, synonymes, variantes et abréviations, de la partie anglaise de la fiche. En fait, ces deux index correspondent aux clés de recherche utilisées lorsqu'on interroge TERMIUM en direct.

La version DOC de TERMIUM contient toutefois deux autres index qui répertorient chaque mot inscrit dans un champ linguistique de la fiche. Ainsi, les troisième et quatrième index comprennent tous les mots (l'un en anglais et l'autre en français), apparaissant dans un syntagme ou une définition. Le terme *disque optique compact* est ainsi fragmenté et indexé sous les rubriques *disque*, *optique* et *compact*.

Par exemple, certains termes peuvent apparaître sur des centaines de fiches. Prenons le mot *disque*. On peut restreindre la recherche à l'aide d'opérateurs logiques, en ajoutant le mot *optique*. Le système cherche donc les fiches où apparaissent les mots *disque* et *optique*, sans qu'ils soient nécessairement juxtaposés. On peut spécifier davantage en ajoutant le mot *compact*, ce qui restreint encore plus le nombre de fiches. Ce mode de recherche donne accès à des fiches qui, même si elles ne donnent pas l'équivalence exacte demandée, peuvent fournir des éléments de recherche ou permettre d'orienter la recherche.

Un cinquième index, celui des domaines, viendra s'ajouter lors de la gravure du prochain DOC. Il sera alors possible de restreindre le nombre d'occurrences de façon encore plus précise.

Dans la même veine, d'autres modifications seront sûrement apportées aux nouvelles versions de notre banque, puisque c'est à l'aide d'outils comme TERMIUM que le Secrétariat d'État peut répondre aux besoins toujours croissants de ses rédacteurs et contribuer à l'épanouissement du français dans les services du gouvernement et dans la société canadienne en général.