Document generated on 07/17/2025 7:20 p.m.

Meta
Journal des traducteurs
Translators' Journal

META

Recherche d'utilisation optimale d'un SGBD en traduction avec aides informatiques : le dictionnaire informatisé bilingue d'écologie des eaux continentales (DIBEEC)

Philippe Thoiron and Dominique Maniez

Volume 34, Number 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/004487ar
DOI: https://doi.org/10.7202/004487ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Thoiron, P. & Maniez, D. (1989). Recherche d'utilisation optimale d'un SGBD en traduction avec aides informatiques : le dictionnaire informatisé bilingue d'écologie des eaux continentales (DIBEEC). *Meta*, *34*(3), 509–515. https://doi.org/10.7202/004487ar

Tous droits réservés © Les Presses de l'Université de Montréal, 1989

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

RECHERCHE D'UTILISATION OPTIMALE D'UN SGBD EN TRADUCTION AVEC AIDES INFORMATIQUES : LE DICTIONNAIRE INFORMATISÉ BILINGUE D'ÉCOLOGIE DES EAUX CONTINENTALES (DIBEEC)

PHILIPPE THOIRON et DOMINIQUE MANIEZ
Centre de Recherche en Terminologie et Traduction
Université Lumière-Lyon 2, France

I. LE DICTIONNAIRE INFORMATISÉ DANS LA PRATIQUE DU TRADUCTEUR

Depuis la généralisation de la pratique du traitement de texte dans la profession, les traducteurs ont pris conscience de l'intérêt de la micro-informatique pour l'élaboration d'outils dictionnairiques complémentaires disponibles en ligne. La diffusion rapide des machines à disque dur a permis de lever les obstacles qui interdisaient tout progrès réel dans ce domaine : la taille des fichiers à stocker et le temps d'accès disque. On est donc arrivé à un stade où le matériel disponible sur le marché permet au traducteur de travailler simultanément avec un traitement de texte et un, ou plusieurs, dictionnaires informatisés. Il reste alors à résoudre les problèmes de logiciels bien sûr, mais il convient aussi de s'interroger sur le contenu du dictionnaire et sur la qualité de ce contenu.

La pratique du traducteur scientifique et technique s'organise autour de deux axes principaux : tout d'abord la maîtrise de la terminologie du domaine, puis la mise en discours de cette terminologie. Pour accomplir cette double tâche, les outils mis à sa disposition sont finalement assez décevants. Les dictionnaires bilingues ne sont bien souvent que des listes d'équivalences terme à terme, où l'on trouve peu de définitions, peu de contextes, et où la présence d'informations concernant l'emploi du terme et ses contraintes collocationnelles constitue l'exception.

Le dictionnaire classique est insuffisant aussi parce qu'il est incapable de suivre le rythme d'évolution des sciences et des techniques. L'arrivée de l'informatique dans le travail du traducteur a rendu évident le fait qu'un outil dictionnairique informatisé pouvait et devait être évolutif. D'abord pour les raisons évoquées ci-dessus, mais aussi parce que les besoins d'un traducteur donné lorsqu'ils sont très spécialisés ne peuvent pas être tous couverts par les outils dictionnairiques disponibles sur le marché. Pour que le dictionnaire soit évolutif, on propose au tradcuteur une structure vide ou partiellement remplie qu'il devra lui-même compléter et mettre à jour[1].

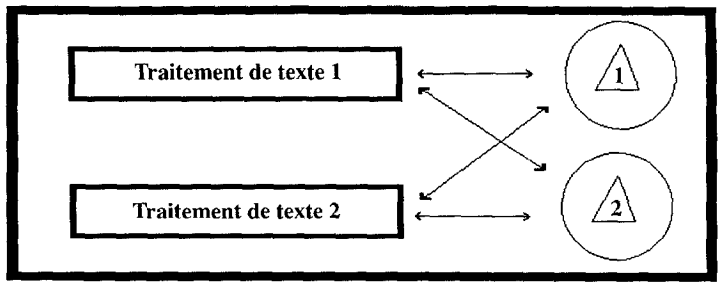
Un autre avantage de l'informatisation est cependant trop souvent négligé. Il s'agit du fait que la structure de l'outil dictionnairique informatisé est beaucoup plus souple que celle du dictionnaire traditionnel, qu'elle est elle-même évolutive. On ne doit donc pas se contenter de reproduire sur un support magnétique la structure de l'ouvrage imprimé mais profiter au contraire des possibilités techniques de consultation rapide de gros fichiers pour imaginer d'autres outils pour le traducteur. Il faut en particulier permettre l'enregistrement et l'exploitation des informations pertinentes sans limitation de longueur.

Lorsqu'il consulte un dictionnaire, un traducteur peut avoir des exigences variées: simple vérification orthographique, confirmation de l'existence d'une équivalence, recherche d'une définition, relation avec les termes voisins, demande de précisions quant à la mise en discours d'un terme, etc. Un outil efficace doit contenir toutes ces informations mais il ne doit pas les imposer à la consultation. Il faut donc envisager une structure des rubriques qui permette, pour un terme donné, une consultation à la carte où les choix pourront éventuellement être hiérarchisés.

Quand les caractéristiques purement terminographiques et linguistiques de cet outil sont arrêtées, et puisque la puissance du matériel disponible est maintenant suffisante, il reste à élaborer les logiciels adéquats. Les pionniers dans ce domaine ont été les traducteurs eux-mêmes qui ont fabriqué leurs propres glossaires bilingues en utilisant par exemple les premières versions de dBase. Il s'agissait bien souvent, et à un double titre, d'un dictionnaire maison: d'abord parce que ce n'était pas l'œuvre d'un lexicographe professionnel, ensuite parce qu'il couvrait un secteur d'activité technique ou scientifique très pointu, voire limité à la pratique et aux préoccupations d'une seule entreprise ou d'un seul laboratoire.

On est alors dans un système où le logiciel de traitement de texte et ses fichiers, et les dictionnaires spécialisés, sont stockés sur un même disque dur mais où le passage des uns aux autres est souvent problématique. La nécessité de programmes d'interface entre, notamment, les traitements de texte et les dictionnaires devient évidente mais leur réalisation est manifestement hors de portée des seuls traducteurs. La situation est évidemment compliquée par la multiplicité des logiciels utilisés, aussi bien pour les traitements de textes (Word, Wordperfect, Wordstar, Textor, etc.) que pour les glossaires où l'on va du simple gestionnaire de fichiers aux différents **SGBD**. La situation peut être représentée par le schéma 1.

SCHÉMA 1



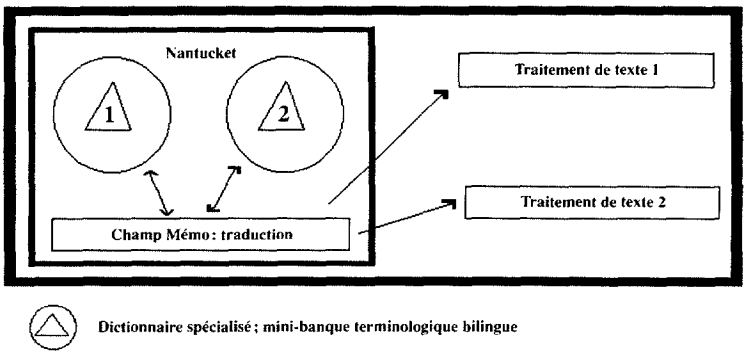
Dictionnaire spécialisé ; mini-banque terminologique bilingue

Les problèmes ont été néanmoins résolus avec l'aide d'informaticiens et les solutions existent. Il reste à les éprouver.

La solution que nous proposons avec le DIBEEC repose sur une stratégie différente. Compte tenu du rôle de standard joué par dBase III Plus dans le domaine des SGBD, il nous a semblé opportun d'explorer les possibilités de ce logiciel et de ses clones non seulement pour constituer le dictionnaire (ce qui est banal) mais pour contribuer à l'élaboration du document traduit lui-même. C'est l'existence des champs mémo qui constitue, comme on le verra, la clé de voûte du système. L'éditeur de texte de ces champs permet en effet au traducteur d'ébaucher et de modifier son texte sans quitter l'environnement dBase. Ce n'est qu'en fin d'opération qu'on aura recours à un véritable logiciel de traite-

ment de texte pour affiner le travail et en achever la mise au net. On se trouve dans la situation représentée par le schéma 2.

SCHÉMA 2



Si la présence des champs constitue pour dBase III Plus un avantage indéniable puisqu'elle libère de la contrainte des champs à longueur fixe et qu'elle autorise la rédaction directe du document traduit, elle ne permet pas de résoudre tous les problèmes. Il faut savoir, en particulier, qu'il est impossible de transférer le contenu d'un champ mémo de dBase III Plus dans une variable chaîne de caractères, ce qui rend impossible tout travail à partir d'un fragment de champ mémo. On ne peut pas, par exemple, programmer une recherche des occurrences d'un terme à l'intérieur d'un champ mémo. Il s'agit, à l'évidence, d'un vice rédhibitoire de ce logiciel pour l'utlisation que nous souhaitons en faire.

On sait d'autre part que le langage de programmation de dBase III Plus est interprété et on peut donc craindre une relative lenteur d'exécution.

Le type de logiciel dont nous avons besoin doit donc, en définitive :

1. utiliser le standard dBase III Plus (notamment à cause de l'intérêt des champs mémo et à cause de la portabilité du produit) ;
2. permettre la récupération et le traitement des champs mémo ;
3. fonctionner en langage compilé pour une plus grande vitesse d'exécution.

C'est parce qu'il remplit ces conditions que nous avons décidé d'utliser Nantucket[2].

II. PRÉSENTATION DU DICTIONNAIRE INFORMATISÉ BILINGUE D'ÉCOLOGIE DES EAUX CONTINENTALES (DIBEEC)

1. STRUCTURE DU DICTIONNAIRE

La macrostructure du dictionnaire est constituée par un ensemble de fiches bilingues (un terme par fiche) contenant chacune plusieurs champs de longueur fixe, et six champs mémo. La fiche est organisée en deux parties totalement symétriques correspondant à chaque langue. Les champs contiennent des informations de divers types :

- documentaire (par exemple source, nom d'auteur de la fiche, date, etc.) ;
- terminologique (par exemple définition) ;
- linguistique (par exemple marques d'usage et contextes).

Définitions, contextes et remarques sont rangés dans les champs mémo. S'agissant des contextes, chaque fiche contient, pour chaque langue, au moins un contexte d'occurrence du terme. S'il y a plusieurs contextes, ceux-ci sont hiérarchisés en fonction de leur intérêt pour le traducteur, c'est-à-dire le plus souvent en fonction de leur valeur colloca-

tionnelle. Seul le premier contexte est référencé dans le champ spécifique «source», les autres sont repérés dans le champ mémo.

2. CONSTITUTION DU DICTIONNAIRE

Un module de saisie a été élaboré et des procédures de correction d'erreur ont été ajoutées par programme aux fonctions standard du logiciel. On peut notamment transférer très simplement le contenu d'un champ mémo dans un autre, évitant ainsi une répétition de la saisie. C'est qu'en effet notre décision de refuser toute limitation de longueur pour la saisie des informations jugées pertinentes (notamment pour les définitions et les contextes) a rendu indispensable le recours aux champs mémo dont certains sont d'ailleurs copieusement documentés.

L'affichage de chaque fiche se fait sur quatre pages-écrans, deux par langue, (cf. schéma 3 et schéma 4).

SCHÉMA 3

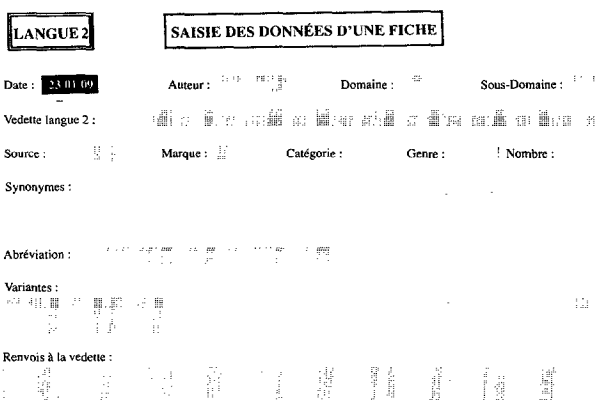
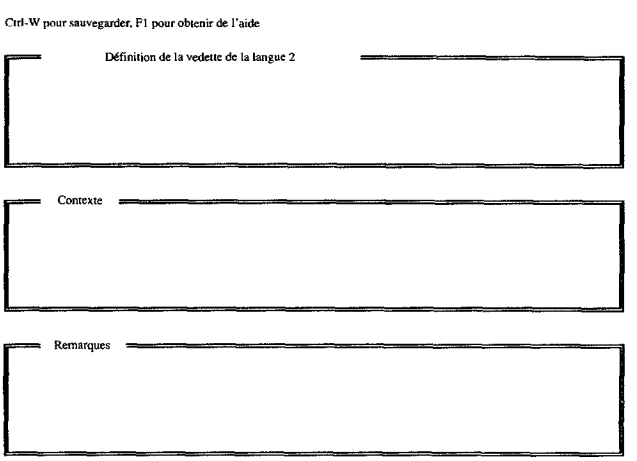


SCHÉMA 4



3. CONSULTATION DU DICTIONNAIRE

Cette opération peut se faire à plusieurs niveaux de «finesse» et à partir de plusieurs stimuli. On peut consulter les seuls champs vedettes L1 et L2 (accompagnées d'informations linguistiques minimales : genre, nombre par exemple) ou y ajouter les définitions ou enfin, obtenir la totalité de la fiche. La consultation multi-niveaux est un auxiliaire précieux dans le cadre de la traduction «fine» puisqu'elle donne accès, lorsque c'est nécessaire, à des emplois terminologiques attestés. Elle incite d'autre part à une bonne gestion du champ «contexte» et notamment à un renseignement soigné de cette rubrique au moment de la constitution du dictionnaire.

Divers stimuli peuvent être utilisés. Il s'agit notamment d'un formant de la vedette (en L1 ou en L2), d'une abréviation, d'un synonyme, mais aussi de tout mot présent dans une fiche (champ mémo compris).

La combinaison de ces deux types de critères permet par exemple l'affichage des vedettes et des définitions d'une fiche dont la vedette française contient le formant «thermo».

On peut aussi accéder à une fiche soit directement par son numéro d'ordre, soit après consultation d'une liste alphabétique des vedettes établie dans l'une des deux langues. Cette liste peut être limitée aux vedettes commençant par une séquence de caractères donnée (par exemple toutes les vedettes commençant par **TRI**). Une telle procédure permet de repérer facilement certaines parentés morphologiques.

Le temps d'accès à une fiche pour une recherche par mot est inférieur à une seconde même si le dictionnaire contient plusieurs milliers de fiches.

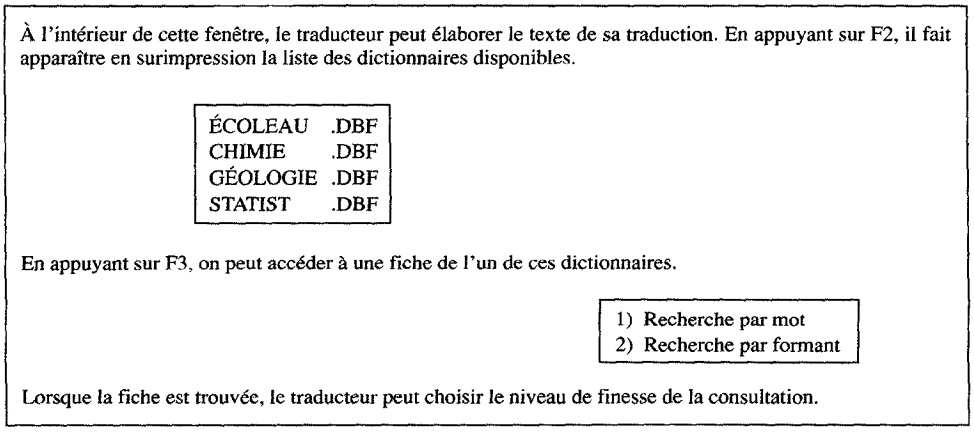
4. UTILISATION SIMULTANÉE DU DICTIONNAIRE ET DE L'ÉDITEUR DE TEXTE

Le traducteur doit pouvoir saisir le texte de sa traduction et consulter, sans changer de logiciel, un dictionnaire spécialisé, voire plusieurs, pour y chercher l'information désirée, au niveau de précision souhaité.

L'utilisation d'un SGBD en langage compilé et muni de champs mémo rend ceci possible dans des conditions de rapidité tout à fait convenables. Le texte de la traduction est ainsi élaboré à l'intérieur même du SGBD. Il s'agit en fait d'une potentialité du système qu'il serait dommage de négliger. Le traducteur a la possibilité de passer d'un dictionnaire à l'autre et d'un degré de finesse à un autre en utilisant les touches de fonction préalablement reprogrammées (cf. schéma 5).

SCHÉMA 5

F1 : Aide F2 : Sélection des dictionnaires F3 : Appel d'une fiche



5. ÉVALUATION DU DIBEEC

Le DIBEEC ne peut pas être représenté comme un produit achevé. Ce qui existe est une ébauche que nous continuons à modifier et pour laquelle le travail de recherche terminologique, que nous avons délibérément passé sous silence ici, est toujours en cours. La structure du dictionnaire dans sa forme actuelle présente les avantages suivants :

- grande portabilité et facilité d'importation et d'exportation de fichiers grâce au standard dBase III Plus dont la notoriété est un gage en ce qui concerne l'évolution et la maintenance du produit ;
- utilisation d'un produit compilé n'exigeant pas l'acquisition de dBase III Plus ;
- possibilité d'exploitation maximale des champs mémo (grâce à la spécificité de Nantucket) aussi bien pour la constitution et l'exploitation du dictionaire que pour l'élaboration du texte de la traduction au sein même du système ;
- vitesse d'exécution ;
- indépendance quasi-totale vis-à-vis des logiciels de traitements de texte puisque la plupart acceptent les fichiers au format dBase ;
- présence d'informations à la fois terminologiques et linguistiques (notamment collocationnelles par référence aux contextes) ;
- possibilité de consultation graduée adaptée au degré de finesse recherché par le traducteur.

CONCLUSION

La possibilité d'élaborer le texte de la traduction au sein même du système n'est évidemment pas la raison d'être du DIBEEC en tant que dictionnaire. C'est néanmoins une caractéristique de ce système qui peut rendre de grands services. En revanche, le concept de consultation multi-niveaux est essentiel à notre projet. Fondamentalement souhaitable et rendu opératoire grâce à l'outil informatique, ce concept est d'autant plus intéressant qu'il conduira très probablement le traducteur à une exigence croissante envers la qualité du contenu des rubriques manifestant, une fois de plus, qu'une avancée technique ne *crée* pas un besoin mais qu'elle le *révèle* en donnant les moyens de le satisfaire.

ANNEXE TECHNIQUE

Nantucket est un compilateur dont le langage de programmation est dérivé du standard dBASE III Plus. Il se distingue du modèle original par les possibilités suivantes:

Nombre d'enregistrements de la base de données : 1 milliard

Nombre de champs par fichier de base de données : 2048

Nombre de fichiers ouverts au même moment : 255 (sous DOS 3,3)

Chaînes de caractères pouvant contenir jusqu'à 64 K octets

Champs mémo traités comme des chaînes de caractères

Exécution dix à vingt fois plus rapide

Le programme s'exécute sur tout micro-ordinateur compatible IBM PC-XT ou PC-AT muni de 640 KOctets de mémoire centrale et tournant sous MS-DOS 3,3 ou ultérieur.

NOTES

1. On observera, au passage, que le traducteur spécialisé doit non seulement fournir une traduction (ce qui est assez normal) mais qu'il doit au préalable fabriquer son outil de travail. Le corollaire de ceci étant que la qualité de la traduction produite est tributaire des aptitudes terminographiques du traducteur. On ne demande pas au médecin de fabriquer son stéthoscope.
2. Pour une description plus complète, voir ANNEXE TECHNIQUE.

Le travail présenté ici, qui est le résultat de la collaboration de trois équipes de recherche (C.R.T.T. Université Lumière Lyon II, Laboratoire d'Écologie des Eaux Douces, Université Claude Bernard Lyon I, GRESLET, Université de Montréal), concerne un outil dont la finalité est double puisqu'il sert la terminologie et la traduction. C'est à ce dernier domaine que nous nous intéresserons plus spécialement en nous attachant à tenir compte d'un état de fait dans le mode de travail du traducteur afin d'en tirer les conséquences utiles.

STRUCTURE DU DIBEEC

N°	Champ	Type	Longueur
1	DATE	Date	8
2	AUTEUR	Caractère	10
3	DOMAINE	Caractère	3
4	SOUS_DOM	Caractère	3
5	VEDETTE_L2	Caractère	60
6	SOURCE_L2	Caractère	8
7	MARQUE_L2	Caractère	3
8	CATGRAM_L2	Caractère	3
9	GENRE_L2	Caractère	3
10	NOMBRE_L2	Caractère	3
11	DEFINIT_L2	Mémo	10
12	CONTEXT_L2	Mémo	10
13	SYNONYM_L2	Caractère	198
14	ABREV_L2	Caractère	30
15	VARORTH_L2	Caractère	160
16	REMARQU_L2	Mémo	10
17	RENVOIS_L2	Caractère	160
18	VEDETTE_L1	Caractère	60
19	SOURCE_L1	Caractère	8
20	MARQUE_L1	Caractère	3
21	CATGRAM_L1	Caractère	3
22	GENRE_L1	Caractère	3
23	NOMBRE_L1	Caractère	3
24	DEFINIT_L1	Mémo	10
25	CONTEXT_L1	Mémo	10
26	SYNONYM_L1	Caractère	198
27	ABREV_L1	Caractère	30
28	VARORTH_L1	Caractère	160
29	REMARQU_L1	Mémo	10
30	RENVOIS_L1	Caractère	160