

Facteurs financiers dans la mise au point d'une banque de terminologie

Banques de terminologie développées comme aide à la traduction automatique basée sur la connaissance

Christian Galinski

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/002788ar>

DOI : <https://doi.org/10.7202/002788ar>

[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

Galinski, C. (1989). Facteurs financiers dans la mise au point d'une banque de terminologie : banques de terminologie développées comme aide à la traduction automatique basée sur la connaissance. *Meta*, 34(3), 567-577.
<https://doi.org/10.7202/002788ar>

FACTEURS FINANCIERS DANS LA MISE AU POINT D'UNE BANQUE DE TERMINOLOGIE

◆ BANQUES DE TERMINOLOGIE DÉVELOPÉES COMME AIDE À LA
TRADUCTION AUTOMATIQUE BASÉE SUR LA CONNAISSANCE ◆

CHRISTIAN GALINSKI
Infoterm
Vienne, Austria

1. INTRODUCTION

Les systèmes de banques de données terminologiques (BDT) représentent un système d'information de type particulier traitant des informations terminologiques, c'est-à-dire un type spécial d'information factuelle sur des notions et des termes (ou autres symboles représentant ces notions), ainsi que des informations sur les sources des informations terminologiques.

Par conséquent, les systèmes développés de BDT sont des systèmes informatisés de documentation terminologique réunissant la terminographie largement informatisée et un système de documentation spécialisé, devenant ainsi une banque de données terminologiques de connaissances (BDTC) traitant la connaissance au niveau de la logique de premier ordre (soit de la logique notionnelle, Budin *et al.* 1988).

Cette étude vise à estimer les coûts de l'établissement d'une banque de données terminologiques et ceux de l'entrée des données terminologiques par rapport aux coûts de la gestion et de la mise à jour de ces dernières.

2. TYPES D'INFORMATIONS TRAITÉES PAR LES BDT

Les notions des différents domaines scientifiques et techniques n'existent pas individuellement mais elles sont créées par l'homme. Les notions nouvelles — dans le cas où elles sont appropriées — se basent toujours sur des connaissances existantes, c'est-à-dire sur d'autres notions. Une fois créé, un symbole existant ou nouvellement construit (soit un terme, soit un autre symbole) est attribué à la nouvelle notion afin de la rendre communicable sous forme parlée ou écrite ou par une représentation graphique ou autre. Par conséquent, les terminologies (notions et systèmes de notions ainsi que les termes et les autres symboles) peuvent être considérées comme des langues artificielles réglées par leurs propres principes qui diffèrent dans une grande mesure de ceux de la langue naturelle. C'est pourquoi la manipulation et le traitement de terminologies exigent des méthodes et des outils particuliers appelés banques de données terminologiques.

Dans une banque de données terminologiques avancée, les notions sont les objets primaires de la gestion de données. Les notions peuvent être représentées de différentes manières par des termes sous différentes formes lexicales, dans un texte et/ou par n'importe quel genre de symbole (p. ex. notations, graphiques, etc.). Le lien entre ces représentations est la notion. Étant donné que le rythme et la manière de développement des termes et des autres symboles diffèrent de celles des notions qui évoluent dans le

cadre de la connaissance, la dynamique des changements notionnels ne peut être contrôlée qu'en enregistrant aussi les données relatives aux systèmes de notions. C'est seulement de cette façon qu'il est possible de réduire la diversité des représentations d'une notion et d'éliminer l'ambiguïté de cette notion pendant la recherche dans le système. Toutefois, le plus grand problème est de savoir comment obtenir des spécialistes de domaine des données suffisamment complètes et fiables pour chaque notion enregistrée et comment les mettre à jour régulièrement tout en assurant leur restructuration automatique.

À présent, on a besoin du cerveau humain pour délimiter les notions de l'une par rapport à l'autre et cela, par des efforts conscients. Pour l'ordinateur, on a besoin d'une banque de données terminologiques plutôt complète et très performante pour identifier les termes (représentant des notions que la banque contient déjà) ainsi que pour de «nouveaux» termes auxquels le système est confronté pour la première fois.

2.1 DÉVELOPPEMENTS DIVERGENTS DES NOTIONS ET DES TERMES

Il n'y a pas de doute que l'évolution de la connaissance est plus rapide et ses changements sont plus fondamentaux que ceux de la langue. En dépit de toute «créativité de langue», la «langue» de Saussure est, à chacun de ses niveaux grammaticaux, régie par la «norme linguistique», ce qui vaut probablement moins pour les langues de spécialités que pour la langue commune. La création de termes nouveaux est entravée en premier lieu, par le nombre très limité d'éléments de termes disponibles pour dénommer des notions nouvelles. Par contre, le nombre des notions augmente à peu près dans la même mesure que le volume des connaissances.

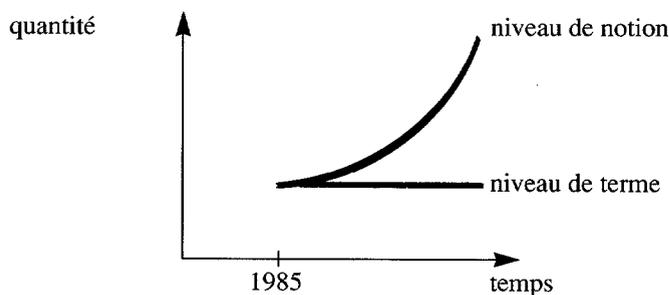


Fig. 1

C'est pourquoi, le développement quantitatif des notions diverge sensiblement de celui des termes. Quelques notions disparaissent tandis que la plupart des nouvelles notions subissent des changements et elles engendrent souvent des systèmes complètement nouveaux. Cette dynamique notionnelle ne peut pas être reflétée ou représentée de manière égale par les termes qui — en tant que symboles linguistiques — témoignent d'une stabilité plus grande que les notions qu'elles expriment. C'est peut-être une des raisons pourquoi une acquisition automatique de la connaissance au moyen d'une identification fiable des termes (découpage des termes) dans des textes spécialisés a échoué jusqu'à présent.

En général, on avance plusieurs arguments en faveur des BDT de type dictionnaire ou même en faveur des banques de termes lexicographiques moins complexes. On

allègue qu'elles n'exigent pas nécessairement une expérience spéciale ni du terminologue ni du terminographe (pour assurer la gestion du système de BDT), ni du spécialiste de domaine (quant à la qualité des données terminologiques préparées pour l'entrée). La conséquence inévitable qu'on doit accepter en est cependant la renonciation à la compétence des deux genres d'experts, ce qui entraîne un niveau proportionnellement bas de performance et de fiabilité.

À présent, le nombre des catégories de données linguistiques-lexicographiques d'entrées de termes figurant dans les dictionnaires pour la traduction automatique dépasse de beaucoup celui des catégories de données terminologiques d'entrées de notions. Les spécialistes de la traduction automatique ont tendance à être absorbés par les problèmes apparaissant à la surface linguistique, c'est-à-dire au niveau du terme des textes spécialisés devant être analysés et traduits par l'ordinateur.

Étant donné cependant que la dynamique du développement des notions ne peut pas être maîtrisée au moyen de termes, il est à conseiller aussi aux experts qui développent des systèmes de traduction automatique de tirer des conclusions à partir de l'expérience qu'on a faite au début du siècle dans le domaine de la terminographie conventionnelle. Au premier quart du siècle, plusieurs projets importants visant à l'établissement de dictionnaires techniques ont échoué parce qu'on n'a appliqué que des méthodes lexicographiques. Par contre, des méthodes terminologiques qui prennent pour point de départ la notion en tant qu'unité de base de la connaissance ont abouti à des résultats qui peuvent être considérés, aujourd'hui encore, comme des modèles. Il est fort probable que des banques de données terminologiques orientées vers la traduction automatique ne manqueront pas de faire leur apparition dans un avenir proche. Étant donné que ces banques contiennent aussi la plupart des données notionnelles (ou peut-être plus encore) provenant du travail terminologique conventionnel accompli par des spécialistes de domaine, il semble être plus sage d'unir les forces au lieu de doubler les efforts en vue de réaliser des données terminologiques de haute qualité pouvant être utilisées pour la traduction automatique, pour la lexicographie spécialisée tout comme pour les systèmes dits de connaissances terminologiques qui servent à des buts différents.

2.2 NOTIONS LINGUISTIQUES GÉNÉRALES ET NOTIONS TECHNICO-SCIENTIFIQUES

Pour ne négliger aucun aspect, il convient de mentionner encore qu'il existe un type de notion représentant des objets concrets ou les phénomènes appartenant à l'environnement de l'homme comme le soleil, la lune, les océans, les lacs, les montagnes, masculin, féminin, etc.. qui font partie intégrale de la langue commune. Ces notions diffèrent de celles de la science et de la technique. Elles peuvent être décrites mais non pas strictement définies ou délimitées. Elles peuvent être regroupées du point de vue de la sémantique plutôt que dans des systèmes de notions.

Il existe probablement d'autres types de notions, particulièrement dans les sciences humaines et sociales, qui reflètent d'ores et déjà quelques traits caractéristiques de la terminologie technico-scientifique et qui, en même temps, gardent certaines ambiguïtés de la langue commune. Il a été prouvé récemment que les principes et les méthodes de la terminologie et de la terminographie technico-scientifique s'appliquent également, sous réserve des adaptations méthodologiques nécessaires, à ces terminologies (infoterm//CEDEFOP 1987). Ni les terminologies scientifiques et techniques, ni celles des sciences humaines et sociales ne peuvent être contrôlées uniquement par des moyens linguistiques et lexicographiques.

La plupart des notions de sciences naturelles et de la technique constituent les «antipodes sémantiques» des notions de la langue commune sus-mentionnées, notions pour lesquelles l'application des méthodes de la linguistique informatisée peut être suffi-

sante. Des méthodes linguistiques-lexicographiques peuvent, bien entendu, être appliquées également aux termes représentant des notions technico-scientifiques, mais elles sont insuffisantes en l'occurrence, puisqu'il s'agit de systèmes de notions et d'autres moyens de représentation de notions.

2.3 COMBINAISON DE DONNÉES TERMINOGRAPHIQUES ET DOCUMENTAIRES

Les données terminographiques sont, en principe, multifonctionnelles. Elles sont structurées conformément à l'ordre de la connaissance. Cette structure est comparable au langage d'indexation des systèmes d'information.

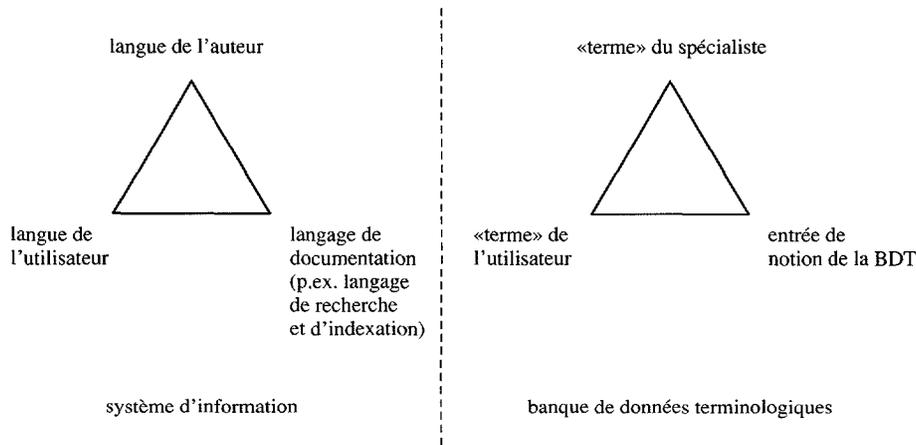


Fig. 2

Les notions (représentées par des termes et/ou d'autres symboles) ainsi que les systèmes de notions peuvent également être utilisés pour la structuration de tout autre type d'information bibliographique ou factuelle. Dans le cas où les BDT et les systèmes d'information sont bien combinés dans une forme plus évoluée, les notions (à l'intérieur du système de notions) représentent la micro-structure de la connaissance (dont une grande partie est contenue p. ex. dans les textes spécialisés), alors que les langages de documentation du système d'information, qui peuvent également être appliqués aux entrées terminologiques, représentent la macro-structure de cette connaissance.

En vue d'assurer

- ◆ une grande précision de la mise à jour et de l'entretien des données terminologiques ;
 - ◆ le contrôle diachronique de la dynamique du développement des notions ;
 - ◆ le contrôle du contenu de textes et de parties de textes ;
 - ◆ le repérage précis à partir des fichiers terminologiques ;
- il est nécessaire
- ◆ d'appliquer les méthodes de la terminographie conventionnelle et informatisée (qui peuvent bien être appliquées en combinaison avec les méthodes lexicographiques informatisées) ;
 - ◆ d'établir les rapports notionnels en plus de l'information qui ne se réfère qu'à la notion en tant qu'unité isolée de la connaissance dont chacune peut être recouverte par une entrée ;

- ◆ d'assurer des liens formels nécessaires entre les entrées représentant les rapports notionnels, ainsi que d'autres rapports en renvois ;
 - ◆ de prévoir des structures d'entrées terminologiques à n dimensions tout comme des structures de fichier à n dimensions *en principe* (permettant, en cas de besoin, une complexité accrue des entrées et des structures du fichier, à une étape ultérieure) ;
 - ◆ de concevoir des interfaces homme-machine faciles à manipuler, nécessaires pour un fonctionnement rapide, sans toutefois limiter, par ces interfaces, la complexité de la banque de données terminologiques ;
- et
- ◆ de combiner les méthodes de la terminographie et de la documentation en vue de maintenir le contrôle sur toutes les sources de l'information terminologique.

Tout cela est valable sans tenir compte de la nature des moyens techniques utilisés, des programmes d'ordinateur exécutés et des applications prévues.

3. TYPES DE BANQUES DE DONNÉES TERMINOLOGIQUES

Les premières grandes BDT ont été des systèmes destinés à soutenir la traduction dans les sections ou services de traduction de grandes compagnies ou d'institutions publiques. Ces BDT contenaient un nombre plus ou moins limité de catégories de données d'un format terminologique conçu à des fins de traduction. Comme les entrées terminologiques que contiennent ces BDT ne sont pas liées pour refléter les rapports notionnels entre les notions, elles représentent des banques de données terminologiques dites de type dictionnaire (BDT/TD) (Felber 1983).

3.1 BDT DE TYPE VOCABULAIRE

Vers le début des années quatre-vingts, ces «grands vieux systèmes» se sont avérés impropres pour l'application par des spécialistes de domaine ayant besoin de l'assistance de l'ordinateur pour leurs activités de recherche et de développement, ainsi que des moyens plus précis pour avoir accès aux informations technico-scientifiques disponibles au lieu de travail ou contenues dans les sources d'information externes. Tout cela nécessite des systèmes de BDT d'une plus grande complexité et plus sophistiqués, basés sur :

- ◆ un format terminologique élargi contenant plus de catégories de données ;
- ◆ une structure d'entrées de grande complexité permettant différentes combinaisons de catégories de données et d'éléments de données, en plus de la possibilité bien connue de répétition par langue et à l'intérieur de la langue (ISO 6156-1987) ;
- ◆ une structure complexe de fichiers terminologiques permettant la représentation de rapports notionnels à l'aide de liens entre les entrées de notions connexes.

Ces genres de systèmes de BDT s'appellent banques de données terminologiques de type vocabulaire (BDT/TV) (Feber 1983). Au cours de l'organisation du premier Congrès international sur la terminologie et l'ingénierie de la connaissance (Czap/Galinski 1987), un système de BDT de type nouveau a été conçu aux fins d'une gestion compréhensive des informations et de la connaissance (Galinski//Nedobity 1987). De tels systèmes sont actuellement en cours de réalisation sous le nom de logistiques de connaissance (Ericsson 1987) et peuvent être appelés banques terminologiques de données de connaissance (BDTC) (Budin et al. 1988).

3.2 BDT DE TYPE DICTIONNAIRE

Le développement technique du matériel et du logiciel a ouvert de nouvelles possibilités pour la conception et l'application de banques de données terminologiques. Dans le passé par exemple, il était à peine possible d'établir des BDT//TV, et cela pour différentes raisons, telles que:

◆ la limitation des configurations du matériel et du logiciel

Quelques-unes des premières banques de termes ont été développées à partir du logiciel prévu pour la documentation, auquel manquait la capacité de répétition ainsi que la possibilité de combinaisons multiples de beaucoup d'éléments de données d'une entrée terminologique ;

◆ la limitation quant aux besoins des usagers et aux exigences de recherche et de sortie

D'autres banques de termes encore ont été conçues à des fins terminologiques ; le nombre des catégories de données a cependant été limité afin d'augmenter la vitesse de recherche et de répondre aux besoins des usagers (pour la plupart des traducteurs) à l'époque de leur établissement ;

◆ le manque d'expérience et de bases théoriques

Beaucoup de catégories de données, en particulier celles destinées à manier les rapports notionnels, n'étaient pas encore suffisamment étudiées ;

◆ le manque de données terminologiques appropriées

Des données terminologiques élaborées systématiquement n'étaient ou bien pas disponibles, ou bien elles n'existaient pas dans les domaines où on en avait besoin ; même des données terminologiques de haute qualité sous forme exploitable sur machines pour l'essai du matériel et du logiciel manquaient ;

◆ le manque de terminologues et de terminographes qualifiés ayant subi une formation adéquate

Il n'existait pas encore de formation pour terminologues ou terminographes conscients de la nécessité et capables de saisir des données terminologiques très complexes.

En général, on n'a pas considéré comme nécessaire de développer des BDT/TV très sophistiquées à des fins qui — apparemment — ne nécessitaient qu'un nombre limité de catégories de données et de fonctions. En outre, on a espéré réduire les coûts et éviter du travail terminologique exigeant beaucoup de temps de la part des spécialistes de domaine. Pour cette raison, on a conçu des systèmes de banques de données terminologiques de type dictionnaire (BDT/TD), d'une complexité et d'une sophistication réduites satisfaisant, comme prévu, en particulier aux besoins limités des traducteurs humains.

C'est une des raisons pourquoi les grands systèmes de BDT existants se sont avérés incompatibles ou inappropriés ainsi pour les systèmes de traduction automatique.

3.3 BANQUES TERMINOLOGIQUES DE DONNÉES DE CONNAISSANCES

Les systèmes terminologiques de connaissances sont développés en tant que composants essentiels d'une gestion compréhensive d'informations et de connaissances, et cela non seulement parce qu'ils sont appropriés pour la mise en mémoire et la recherche, mais aussi parce que nos langues naturelles, techniques et parlées se prêtent mal à la tâche d'ordonner les éléments de la connaissance (y compris des informations terminologiques et d'autres informations factuelles) d'une manière systématique, ainsi que de repérer les textes stockés et d'autres formes de représentation de la connaissance complexe. Dans le cas où un tel système doit continuer à fonctionner d'une manière adéquate après les phases de l'établissement et de la mise en service, en particulier dans un environnement multifonctionnel où la quantité totale des informations augmente sensiblement, il exige, dès le début, un minimum de complexité et de sophistication.

Les banques de données terminologiques avancées décrites plus haut sont, en principe, multifonctionnelles tout comme les données terminographiques qu'elles contiennent. Ces BDT se prêtent à des fins différentes dans le cadre de la gestion et du transfert des informations/connaissances. La traduction automatique profiterait sans doute de l'existence de telles banques de données terminologiques avancées qui — comme les

banques de données de connaissances — pourraient compléter les dictionnaires électroniques nécessaires dans les systèmes de traduction automatique avec les données nécessaires pour la traduction hautement fiable de textes spécialisés.

4. COÛTS DE L'ÉTABLISSEMENT ET DE LA GESTION D'UNE BDT

Dans le cas de quelques grands systèmes de BDT/TD conçus afin d'assister la traduction humaine, on a supposé qu'une banque de données terminologiques d'une structure simple et d'une complexité réduite servirait le mieux les besoins «limités» des usagers (c'est-à-dire les traducteurs). Or, cela n'est valable que dans des conditions très spéciales. Si, par exemple, le nombre des synonymes est radicalement réduit avant l'entrée des données par quelque genre de régulation terminologique, si la diversité des types de texte est limitée, les textes mêmes sont d'une nature très homogène ou même quasi-normalisée et que la banque de données terminologiques est, dans son ensemble, plus ou moins monofonctionnelle, simple ou même primitive, les BDT peuvent s'avérer tout à fait efficaces et satisfaisantes quant aux résultats.

Dans tout autre cas où les BDT sont appliquées d'une manière multifonctionnelle ainsi que dans un environnement où le développement de la connaissance et de sa représentation sémiotique n'est pas limité, une décision en faveur d'un système de BDT simple et peu complexe sera reconnue comme fautive à un stage où elle pourra à peine être corrigée. Tout cela mène à une situation où l'on est contraint soit, d'accepter la diminution permanente de la performance du système et l'accroissement des coûts de l'entretien et de la mise à jour des données, soit de reconcevoir complètement l'entier système de BDT. Dans la plupart des cas, lorsqu'un système est reconçu, seulement un nombre réduit des entrées ou des données terminologiques contenues peuvent être transférées du vieux système au système nouveau à l'aide de programmes de conversion sans intervention humaine.

Si l'on essaie d'économiser sur l'entrée au cours du développement et de la réalisation de la BDT, on doit être conscient du fait, que cette somme devra être remboursée à un taux beaucoup plus élevé pendant l'exploitation de la BDT. Il sera également nécessaire de régler l'interaction usager — système afin d'éviter qu'il devienne insupportablement coûteux. Avec le temps, un tel système de BDT sera loin d'atteindre le but envisagé lors de son établissement et d'accomplir la tâche initialement prévue.

Seule une BDT sans limitations intégrées, permettant un développement ultérieur afin de répondre à des exigences ultérieures, sera capable d'élargir sa capacité de stockage, de s'adapter aux nouveaux développements et de maîtriser la dynamique du développement notionnel.

4.1 RAPPORT ENTRE LE DEGRÉ DE COMPLEXITÉ ET DE SOPHISTICATION ET LES COÛTS DE L'ÉTABLISSEMENT ET DE LA GESTION D'UNE BDT

- ar : dépenses pour le travail humain nécessaires pour l'entretien, la mise à jour et la recherche
- as : dépenses pour le travail humain nécessaire à la mise en mémoire
- * On admet que les dépenses tolérées pour le travail humain augmentent légèrement en fonction de l'importance de la base de données. Ce fait est dû à son utilité croissante.
- ** On admet que l'importance de la base de données et la fréquence des recherches augmentent avec le temps.

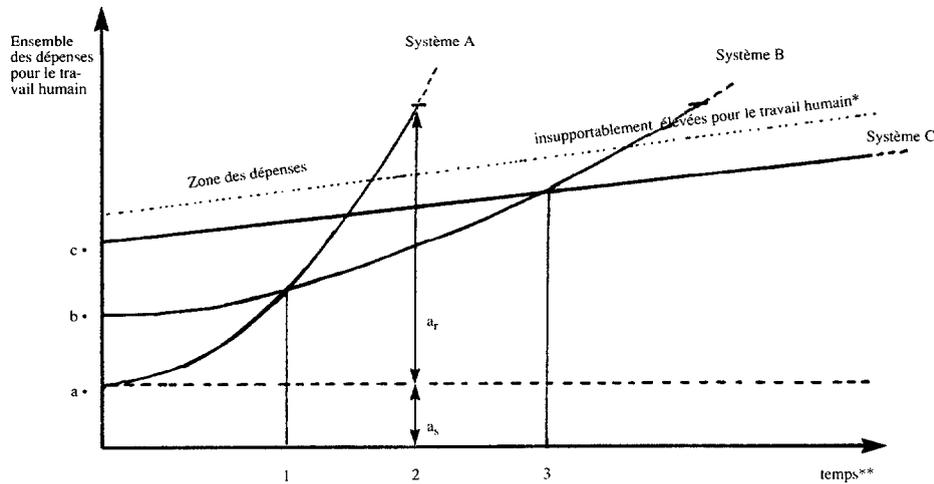


Fig. 3

Cette figure illustre trois systèmes de banque de données terminologiques, notamment «A», «B» et «C».

«A» est un système à complexité et sophistication réduites nécessitant des dépenses initiales basses pour le travail humain et peu d'investissements en matériel et en logiciel, parce qu'à l'époque où ce système est lancé, les coûts de recherche sont nuls et les coûts de mise en mémoire sont relativement bas. Au moment où le système est cependant mis en service, ses coûts montent en flèche, d'où le taux réduit de sa survie.

«B» est un système d'un degré moyen de complexité et de sophistication nécessitant plus de dépenses b pour le travail humain et plus d'investissements en matériel et en logiciel que le système «A». Au début de son fonctionnement, la courbe représentant les coûts ressemble beaucoup à celle du système «A».

Ensuite, il montre ses avantages par rapport au système «A» quant aux coûts de la mise en mémoire, de l'entretien, de la mise à jour et de la recherche. Bien que sa capacité de stockage soit nettement supérieure à celle du système «A», sa force de survie est également limitée.

Par rapport aux systèmes «A» et «B», le système de BDT «C» est beaucoup plus coûteux au début, mais l'ensemble de ses coûts de fonctionnement, qui comprennent ceux de l'enregistrement, de l'entrée, de la mise à jour, de l'entretien et de la recherche de données terminologiques, et qui sont déterminés largement par la quantité du travail humain investi, seront plus bas au bout de quelques années et n'augmenteront pas de manière exponentielle. Ce système a cependant le désavantage qu'à la période de démarrage, il est coûteux et que ses avantages ne se manifestent que plus tard, et en particulier, dans un environnement multifonctionnel ou dans le cas où la quantité des données stockées serait très importante.

Pour toutes ces raisons, les coûts du système «C» doivent être considérés comme un investissement minimum permettant d'atteindre l'objectif d'un système de BDT performant, fiable, multifonctionnel et stable.

Le système de BDT «C» représente une banque de données terminologiques de connaissance (BDTC) traitant toute sorte d'information écrite ou graphique en tant

qu'unités de la connaissance ou parties de textes, tout en utilisant une combinaison de méthodes terminographiques avancées (pour contrôler la «macro-structure» de la connaissance au moyen des langages dits de documentation subdivisant les documents, y compris les enregistrements terminologiques et autres, les données bibliographiques, des résumés ou n'importe quelles parties d'un document, en unités maniables).

4.2 AVANTAGES D'UNE BDT/TV ET D'UNE BDTC

Les rapports notionnels réalisés dans les BDT permettent

- ◆ de clarifier l'équivalence entre les termes de langues différentes représentant la même notion, et cela à un degré de fiabilité plus élevé ;
- ◆ de trouver des termes inconnus pour des notions dont on ne connaît que la position qu'elles occupent dans le système de notions ;
- ◆ de contrôler la synonymie ainsi que l'homonymie (en particulier les séries horizontales et verticales après la révision de n'importe quelle entrée d'homonymes) ;

et

- ◆ une mise à jour/révision au moins semi-automatique des entrées connexes.

Par l'intermédiaire d'une notion et de son rapport notionnel avec d'autres notions, l'utilisateur est conduit vers les unités individuelles de la connaissance et de l'information qu'il cherche. De cette manière, la perte d'information est sensiblement réduite. En même temps, on peut maintenir à un niveau contrôlable, le «bruit» (c'est-à-dire les informations fausses ou ambiguës qui mènent à des réponses hors de propos ou même trompeuses). Par contre, les systèmes BDT/TD exigent des efforts humains croissants afin de réduire le bruit à un niveau tolérable.

Chaque nouvelle entrée où la révision des entrées existantes nécessite le contrôle d'autres entrées — en particulier celles qui sont en rapport notionnel avec elles. Dans le cas où le système ne fournit aucune indication quant aux entrées devant être vérifiées, le temps et les efforts nécessaires pour le contrôle et la vérification des entrées augmentent d'une manière exponentielle si l'on veut maintenir la haute fiabilité, alors que l'efficacité du processus du contrôle diminue.

4.3 ACCEPTATION PAR LES USAGERS ET EXIGENCES QUANT À LA QUALIFICATION DU PERSONNEL

À première vue, les BDT à complexité réduite pourraient paraître sensiblement plus conviviales étant donné qu'elles obviennent à la nécessité d'une instruction pour l'application du système qui prend du temps. Ses désavantages se manifestent cependant sans tarder sous forme d'un niveau élevé de bruit et d'une réduction de fiabilité des éléments individuels d'informations, conduisant à la perte de la fiabilité de l'ensemble du système. En plus de cela, les spécialistes de l'information sont toujours tenus de se servir de ces systèmes d'une manière efficace. Néanmoins, le genre et l'importance de la perte d'informations d'une part et l'augmentation du bruit d'autre part prendront un développement que ni l'utilisateur ni l'opérateur du système ne pourront plus maîtriser.

4.4 MAINTIEN DE LA FIABILITÉ DES DONNÉES

Une BDT à complexité et sophistication réduites ne permet que l'application de méthodes formelles dans l'exécution de l'entretien et de la mise à jour des informations terminologiques. Cela conduit inévitablement à une augmentation exponentielle des efforts humains et, par conséquent, des coûts du maintien de la fiabilité à un niveau constant, même si la croissance de la quantité des informations terminologiques est linéaire.

Afin de permettre l'entretien et la mise à jour automatiques ou au moins semi-automatiques des données, le nombre des catégories de données du format terminologique, la

complexité de la structure du fichier et le nombre des fichiers, ainsi que les liens entre les fichiers doivent être augmentés, alors que la quantité des données terminologiques s'accroît constamment. Il convient donc de concevoir des systèmes de BDT :

- ◆ avec une structure d'entrées à n dimensions permettant un accroissement sans restrictions des catégories de données ;
- ◆ avec une structure de fichiers à n dimensions ;
- ◆ sans restrictions concernant le nombre des fichiers et le genre de liens entre les fichiers ;

en principe — dans la pratique on sera obligé d'accepter des compromis. Sinon, on met en cause la fiabilité des unités individuelles des informations terminologiques qui garantissent la fiabilité générale du système de BDT. Plus les catégories de données, d'éléments de données et d'entrées sont liés mutuellement de façon formelle et notionnelle, plus le système permet la mise à jour et l'entretien systématiques au moins semi-automatiques de son contenu. De cette manière, l'information terminologique sur les nouvelles entrées ainsi que les changements survenus dans les entrées existantes sont comparés immédiatement avec les informations stockées dans d'autres entrées reliées au niveau notionnel ou formel. De cette façon, le «bruit» et la perte d'informations sont réduits alors que le niveau élevé de la fiabilité de l'ensemble du système est maintenu.

5. L'EFFET DE RESTRICTIONS SUR LE DOMAINE D'APPLICATION D'UNE BDT

Comme c'est le cas pour les systèmes de traduction automatique, l'efficacité des BDT peut être augmentée considérablement en limitant rigoureusement leur domaine d'application. Comme résultat de telles restrictions, les banques peuvent être :

- ◆ monofonctionnelles (p.ex. orientées uniquement vers l'aide à la traduction) ou bien multifonctionnelles ;
- ◆ limitées en ce qui concerne:
 - le nombre des domaines de spécialité (p. ex. des nomenclatures scientifiques et techniques)
 - les terminologies harmonisées ou unifiées (y compris la terminologie normalisée)
 - le nombre des langues (p.ex. seulement des langues appartenant à la même famille)
- ◆ conçues spécialement à l'intention d'un groupe particulier d'utilisateurs (ou de quelques-uns très apparentés).

On pourrait alors avoir l'impression qu'un degré faible de complexité et de sophistication suffirait et que les affirmations ci-dessus ne seraient pas pertinentes. Dès que ces restrictions seront cependant supprimées pour utiliser le système :

- ◆ de manière multifonctionnelle ; ou
 - ◆ par de nouveaux groupes d'utilisateurs ; ou
 - ◆ par des langues nouvelles et/ou plus éloignées du point de vue linguistique ; ou
 - ◆ pour des types différents de terminologies (p. ex. des domaines interdisciplinaires) ; ou
 - ◆ pour stocker des grandes quantités de données au lieu d'un nombre restreint d'entrées ;
- il sera impossible, ou du moins très difficile et onéreux, d'adapter la BDT à son nouvel environnement — pourvu qu'un tel développement n'ait pas été prévu au cours de la phase de conception et de spécification du système.

CONCLUSIONS

Si l'on compare la connaissance à un grand magasin avec beaucoup de cloisons et d'étagères stockant beaucoup de conteneurs remplis de parties ou d'éléments des produits, les langages de documentation représentent l'ordre des cloisons, des étagères et des conteneurs, tandis que la terminologie représente l'ordre à l'intérieur des conteneurs. Au cours du développement de produits nouveaux, le degré des changements survenus dans

les conteneurs est plus élevé que celui des conteneurs, des étagères et des cloisons, et en fin de compte, de l'ensemble du magasin.

Le style de direction le plus ancien et le plus répandu est la «gestion par ignorance». Les conséquences, parfois désastreuses de ce style, sont évidentes depuis bien longtemps. De nos jours, la gestion d'informations est reconnue comme ayant une importance vitale pour toute sorte d'institution et d'organisation. Il est pourtant surprenant de voir le peu d'effort fait dans le domaine de l'information pour acquérir les connaissances de base et les expériences relatives aux principes et aux méthodes de la terminologie et aux langages de documentation qui sont essentielles pour la gestion d'informations (comp. GfK-SIG/IS 1985).

La «gestion par ignorance» dans la gestion d'informations de n'importe quelle institution ou organisation mène inévitablement à la «gestion de l'ignorance» de l'ensemble de l'institution ou de l'organisation ayant pour résultat du «bruit», des pertes dans la communication, de la perte de temps lors de la recherche d'informations, etc. ce qui se traduit par des coûts élevés d'exploitation. Tout cela vaut également pour l'utilisation de systèmes de BDT à complexité et sophistication réduites dans la traduction automatique.

RÉFÉRENCES

- BUDIN, G. *et al.* (1987) : «Terminology and Knowledge Data Processing», CZAP, H. ; GALINSKI, C. [eds.] *Terminology and Knowledge Engineering*, Supplement, Proceedings of the International Congress on Terminology and Knowledge Engineering, 29 Sept. — 1 Oct. 1987, Trier. Indeks, Frankfurt.
- CEDEFOP/Infoterm. [eds.] (1987) : *Tools for multilingual institutional work in the field of vocational training*, 1st ed. A CEDEFOP-Infoterm publication, Berlin.
- CZAP, H. ; GALINSKI, C. [eds.] (1987) : Terminology and Knowledge Engineering, Proceedings of the International Congress on Terminology and Knowledge Engineering, 29 Sept. — 1 Oct. 1987, Trier, Frankfurt.
- ERICSSON [ed.] (1987) : *Beiträge zur Wissenslogistik. CAT — Computer aided translation. I. Anwendertreffen*. Stuttgart, Ericsson.
- FELBER, H. (1983) : «Computerized terminography in TermNet : the role of terminological data banks». *Term banks for tomorrow's world, Translation and the Computer 4*. Proceedings of a conference, 11-12 November 1981, London, pp. 8-20.
- GALINSKI, C ; NEDOBITY, W. (1986) : «A terminological data bank as a management tool», *Infoterm*, 9 p., A4, (Infoterm 3-86 en), Wien.
- GESELLSCHAFT FÜR KLASSIFIKATION. SPEZIELLE INTERESSENGRUPPE INDEXIERUNGSSPRACHEN (SIG-IS). [ed.] (1985) : *Recommendations for Classification, EK-03 (en) Free text in information systems. Capabilities and Limitations classification*. EK-03 (en) (GfK/SIG IS), International Classification 12 [2], pp. 95-98.
- ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). [ed.] (1987) : *Format d'échange sur bande magnétique des données terminologiques et/ou lexicographiques (MATER)*, Genève, ISO, 25 p., A4, (ISO 6156 : 1987).