

Les bases de connaissances terminologiques : les banques de terminologie de seconde génération

Gabriel Otman

Volume 42, numéro 2, juin 1997

Lexicologie et terminologie II (1) et Traduction et post-colonialisme en Inde

Translation and Postcolonialism: India (2)

URI : <https://id.erudit.org/iderudit/003772ar>

DOI : <https://doi.org/10.7202/003772ar>

[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

Otman, G. (1997). Les bases de connaissances terminologiques : les banques de terminologie de seconde génération. *Meta*, 42(2), 244–256.
<https://doi.org/10.7202/003772ar>

Résumé de l'article

Une base de connaissance terminologique (BCT) est à la fois une banque de terminologie conceptuellement et sémantiquement structurée et une base de connaissances. Les informations terminologiques, que nous qualifions de connaissances, y sont décrites sous la forme de réseaux conceptuels et sémantiques que nous appelons "réseaux sémantico-terminologiques" (RST). Les RST permettent de représenter, sous la forme de graphes, les unités terminologiques d'un domaine et les relations sémantiques qui les unissent au sein d'un système notionnel. Ces relations, au nombre de six dans notre modèle, possèdent la double fonction de regrouper les notions qui ont un certain degré de parenté pour mieux les distinguer par leurs traits distinctifs au sein de classes et de sous-classes.

LES BASES DE CONNAISSANCES TERMINOLOGIQUES : LES BANQUES DE TERMINOLOGIE DE SECONDE GÉNÉRATION*

GABRIEL OTMAN

Université de Paris XIII, Villetaneuse, France

Résumé

Une base de connaissance terminologique (BCT) est à la fois une banque de terminologie conceptuellement et sémantiquement structurée et une base de connaissances. Les informations terminologiques, que nous qualifions de connaissances, y sont décrites sous la forme de réseaux conceptuels et sémantiques que nous appelons «réseaux sémantico-terminologiques» (RST). Les RST permettent de représenter, sous la forme de graphes, les unités terminologiques d'un domaine et les relations sémantiques qui les unissent au sein d'un système notionnel. Ces relations, au nombre de six dans notre modèle, possèdent la double fonction de regrouper les notions qui ont un certain degré de parenté pour mieux les distinguer par leurs traits distinctifs au sein de classes et de sous-classes.

Abstract

The first generation of terminology banks was developed before and during the eighties. Today we have the know-how to envisage a second generation of terminology banks i.e. Terminological Knowledge Bases (TKB). A TKB is both a knowledge base and a conceptually and semantically structured terminology bank. Terminological knowledge is represented in the form of a semantic network called TSN (Terminological Semantic Network). TKBs are intended to be incorporated into knowledge-based systems developed by knowledge engineers and into workstations used by technical communicators (translators, technical writers, terminologists, documentalists, information systems administrators...).

1. LES BANQUES DE TERMINOLOGIE DE PREMIÈRE GÉNÉRATION

Les années 80 ont vu le développement des banques de terminologies informatisées et l'émergence de la notion de terminotique. Historiquement la création de la première banque de terminologie informatisée remonte à 1963, avec la mise en place à Luxembourg de DICAUTOM qui deviendra EUROIDICAUTOM en 1973. LEXIS est né en Allemagne en 1966 et NORMATERM a été créé par l'AFNOR en France en 1973. Les Canadiens ont successivement créé TERMIUM en 1970 et la BTQ (Banque de Terminologie du Québec) en 1973 (Paradis et Auger 1987). Mais ce n'est qu'au cours de la décennie suivante, avec l'émergence de la micro-informatique et de la télématique, qu'elles ont pris leur plein essor. Bien que leur nombre se soit parallèlement multiplié, les précurseurs cités plus haut restent les banques les plus réputées et les plus consultées. Elles constituent l'ossature des banques de terminologie de la première génération.

Malgré les services réels que ces banques rendent à leurs utilisateurs, ces derniers expriment néanmoins, chaque fois qu'on les consulte, diverses critiques qui vont de l'insatisfaction quand la consultation ne donne pas les résultats escomptés à la frustration quand elle n'aboutit pas pour des raisons informatiques ou télématiques. Lerat (1995) distingue cinq raisons d'insatisfaction : le mode de conception même de ces banques, leur finalité,

leur alimentation, leur mise à jour et leur épuration. Nous en ajouterons une sixième, l'absence de structure multirelationnelle. Détaillons ces divers points. Même si de nombreuses banques sont accessibles au grand public (via CD-Rom ou télématique), elles ont été initialement conçues pour un public restreint et très ciblé : les traducteurs des services de la Communauté européenne à Luxembourg, Bruxelles et Strasbourg pour EURODI-CAUTOM, les langagiers de l'OLF pour la BTQ, du Bureau des traductions pour TERMIUM, les normalisateurs de l'AFNOR pour NORMATERM, le personnel de l'entreprise et ses sous-traitants pour TEAM de Siemens. Les fiches dont ces banques sont constituées sont souvent des extraits de textes accompagnés de références bibliographiques (TERMIUM en particulier), qui comportent très peu d'informations linguistiques, c'est-à-dire grammaticales, sémantiques et phraséologiques. La démarche est cumulative, elle procède par dépouillement essentiellement. Les sources et les informateurs sont multiples. Les risques de redondance et de dispersion de l'information sont conséquents. L'approche adoptée par les concepteurs est donc fortement sémasiologique puisque les informateurs partent des termes pour déterminer les concepts sous-jacents. La taille de ces banques terminologiques, leur richesse informative et l'absence, historiquement parlant, d'outils performants de gestion des données terminologiques sont autant de raisons qui peuvent expliquer qu'elles ne soient que rarement correctement mises à jour et épurées de leurs informations obsolètes ou erronées.

Jusqu'à la fin des années 1980, les utilisateurs de ces banques de données terminologiques se satisfaisaient de ces «grands fichiers prédictionnaires» (Humbley 1994) que l'on consultait essentiellement comme des dictionnaires papier à partir d'une forme alphabétique dans une langue donnée. Tel est encore le principe essentiel d'une consultation sur TERMIUM, par exemple. Des outils de navigation performants de type hypertexte se répandent aujourd'hui. Parallèlement, terminologues, traducteurs et documentalistes ont pris conscience des aspects systématiques et onomasiologiques des terminologies. Une terminologie digne de ce nom est un objet structuré et cette structure doit être visible. Nous ne prétendons pas que les terminologues des années 1960 à 1990 ne dressaient pas de terminologies structurées : nous précisons simplement que dans la très large majorité des banques de terminologies actuelles, cette structure n'est ni visible ni accessible. En d'autres mots, le cheminement de lecture ne peut pas s'effectuer en suivant les lignes de construction de cette structure sous-jacente. Or, il se trouve qu'à notre sens, nous avons atteint aujourd'hui un point de rupture. En effet, les outils et les esprits sont dorénavant prêts pour une nouvelle génération de banques de données terminologiques fondées sur la notion de modèle relationnel. Nous les appellerons base de connaissances terminologiques (BCT).

2. LE MODÈLE RST

Nous proposons dans cette perspective un modèle relationnel inspiré des réseaux sémantiques de Quillian (1968) que nous appelons réseau sémantico-terminologique (RST). Les RST permettent de représenter sous la forme de graphes les unités terminologiques d'un domaine et les relations sémantiques qui les unissent au sein d'un système notionnel. Ces relations, au nombre de six dans notre modèle (*sorte-de*, *partie-de*, *fonction-de*, *proximité-de*, *équivalent-à*, *contraste-avec*), possèdent la double fonction de regrouper les notions qui ont un certain degré de parenté pour mieux les distinguer par leurs traits distinctifs au sein de classes et de sous-classes. Le modèle est présenté dans Otman (1990, 1992a et 1992c) et détaillé dans Otman (1995) où il a fait l'objet d'une préapplication sur la terminologie de l'IA et, en particulier, sur le terme «règle», ses hyponymes et hyperonymes (voir figures 1 et 2). Il a permis de représenter la plus large part des informations contenues dans les définitions de notre dictionnaire de l'IA. Sur le plan des limites, nous

convenons que ce formalisme ne parvient pas à prendre en compte l'expression de modalités et de restrictions qui sont exprimées dans les définitions par des marqueurs ou des formules comme : *le même, dans la plupart des cas, systématiquement, quel que soit, toutes... à l'exception de*. Il faut en conclure qu'il serait fâcheux de faire l'impasse totale sur la définition. Une définition en langue naturelle telle qu'on la trouve dans un dictionnaire terminologique courant reste utile — voire indispensable — dans un système de requêtes électronique. Le formalisme est un moyen de recherche d'informations, de présentation systématisée d'informations d'un certain type et de mise en relation de termes d'un micro-domaine. Il ne remplace pas la richesse expressive de la langue naturelle. Au demeurant, l'adjonction d'une base de données dictionnaire de définitions accessible par l'ouverture d'une fenêtre «définition» associée à chaque nœud ne pose aucun problème informatique avec une interface de type hypertexte, par exemple. Dans notre cas, l'essentiel n'était pas de lever ces deux incohérences — c'est une affaire de questionnement de spécialistes et de recherche de contextes adéquats — mais de souligner qu'un système à base de connaissances construit autour du principe des RST est en mesure d'attirer l'attention du terminologue sur la présence de ces incohérences. Dès lors, ce dernier a le choix entre conserver ces incohérences en connaissance de cause ou les lever après recherches approfondies. Le système de requêtes que nous proposons n'a pas pour vocation d'avoir réponse à tout, il s'agit simplement d'un système d'aide à la constitution de bases terminologiques. À ce titre, il peut également s'avérer productif en informations nouvelles.

Dans le cadre de l'étude que nous menons sur la terminologie de géomorphologie (voir 5.1.), nous avons été amené à dresser les RST des termes «vallée» et «glacier» à partir des définitions et des informations fournies par deux sources : l'une à vocation terminologique, le VDLG (*Vocabulaire de la géomorphologie*, CILF 1979), l'autre à vocation lexicologique, le TLF (*Trésor de la Langue Française*). Dans ce cas également, le modèle a permis de représenter une part importante des informations de départ. Il n'en demeure pas moins que les deux réseaux obtenus (voir figures 3 et 4) sont largement divergents mais individuellement cohérents dans une large mesure. N'est-ce pas là finalement la qualité essentielle d'un dictionnaire spécialisé ou d'une banque de terminologie ?

Hyperonymes	savoir savoir humain sous-ensemble du savoir humain connaissance connaissance opératoire							
Terme de base	REGLE							
Hyponymes de 1er niveau	règle de chaînage	règle pondérée		règle d'inférence	règle démon	règle à niveau	règle de déduction	règle de priorité règle de production règle valide méta-règle
Hyponymes de 2ème niveau	règle en avant règle en arrière	règle incertaine règle précise	règle imprécise règle précise	• règle de modus • spécialisation universelle • hyperrésolution • règle d'inférence saine • règle d'inférence par défaut • règle d'inférence non monotone • règle d'inférence monotone	r. antitélém r. non antitélém	r. de bas niveau • r. de haut niveau	généralisation	
Hyponymes de 3ème niveau	règle antécédente règle autoréférentielle			modus ponens modus tollens				règle de définition

Figure 1 :
Tableau synoptique de l'ensemble des hyponymes et hyperonymes de règles

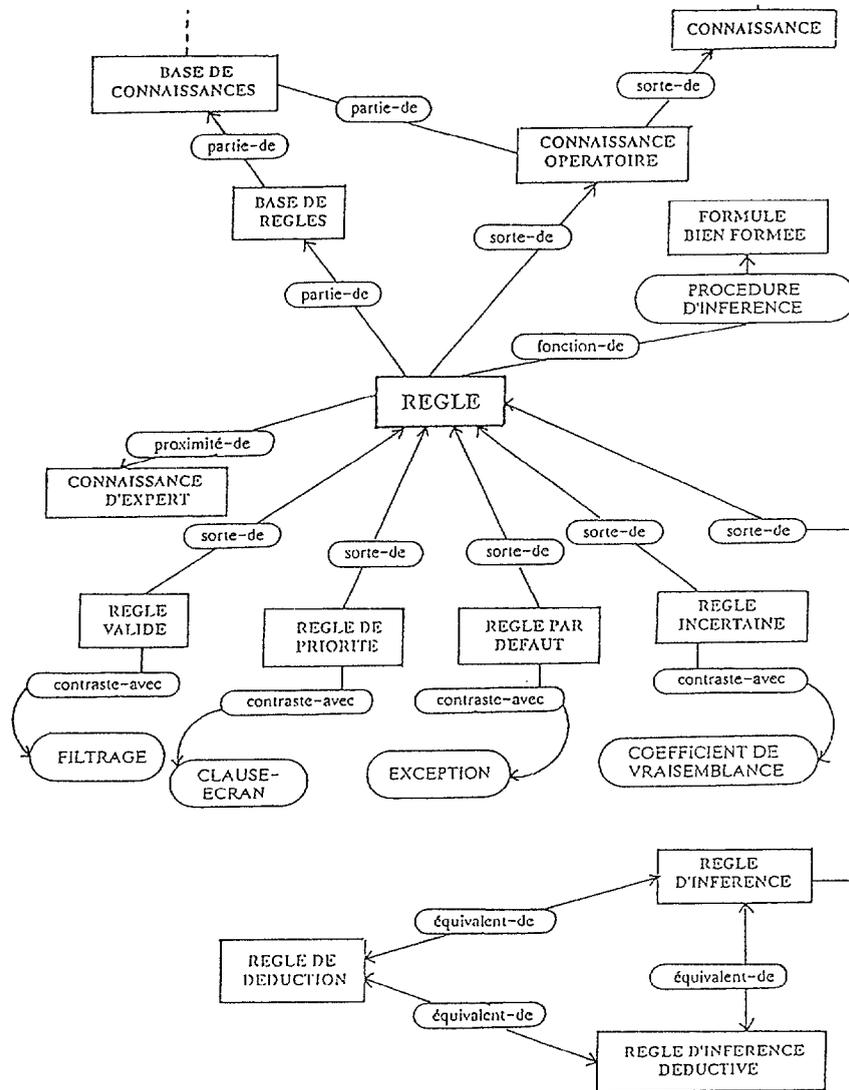


Figure 2 :
Fragment du RST de règle

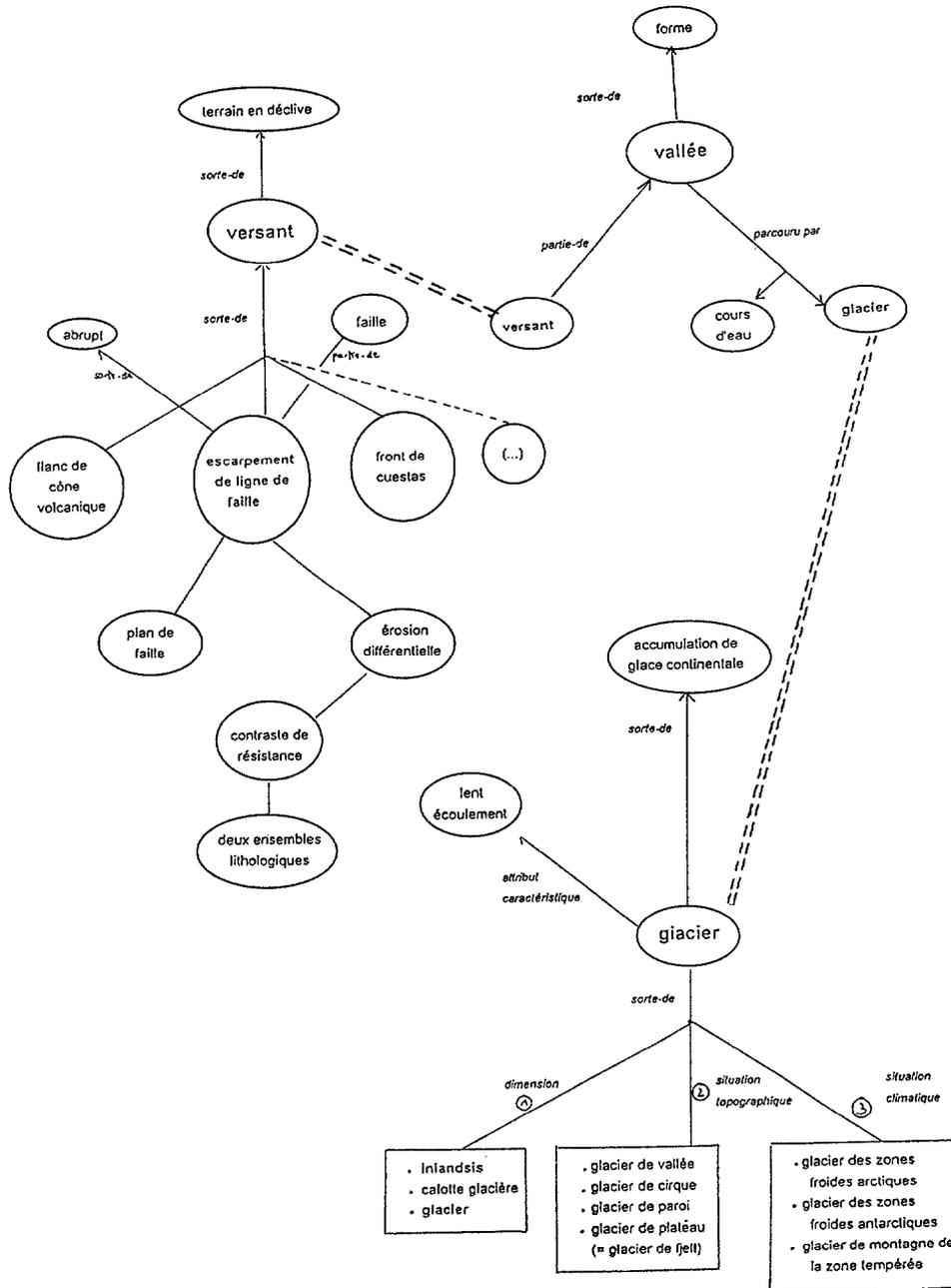


Figure 3 :
RST de vallée et glacier à partir du VDLG

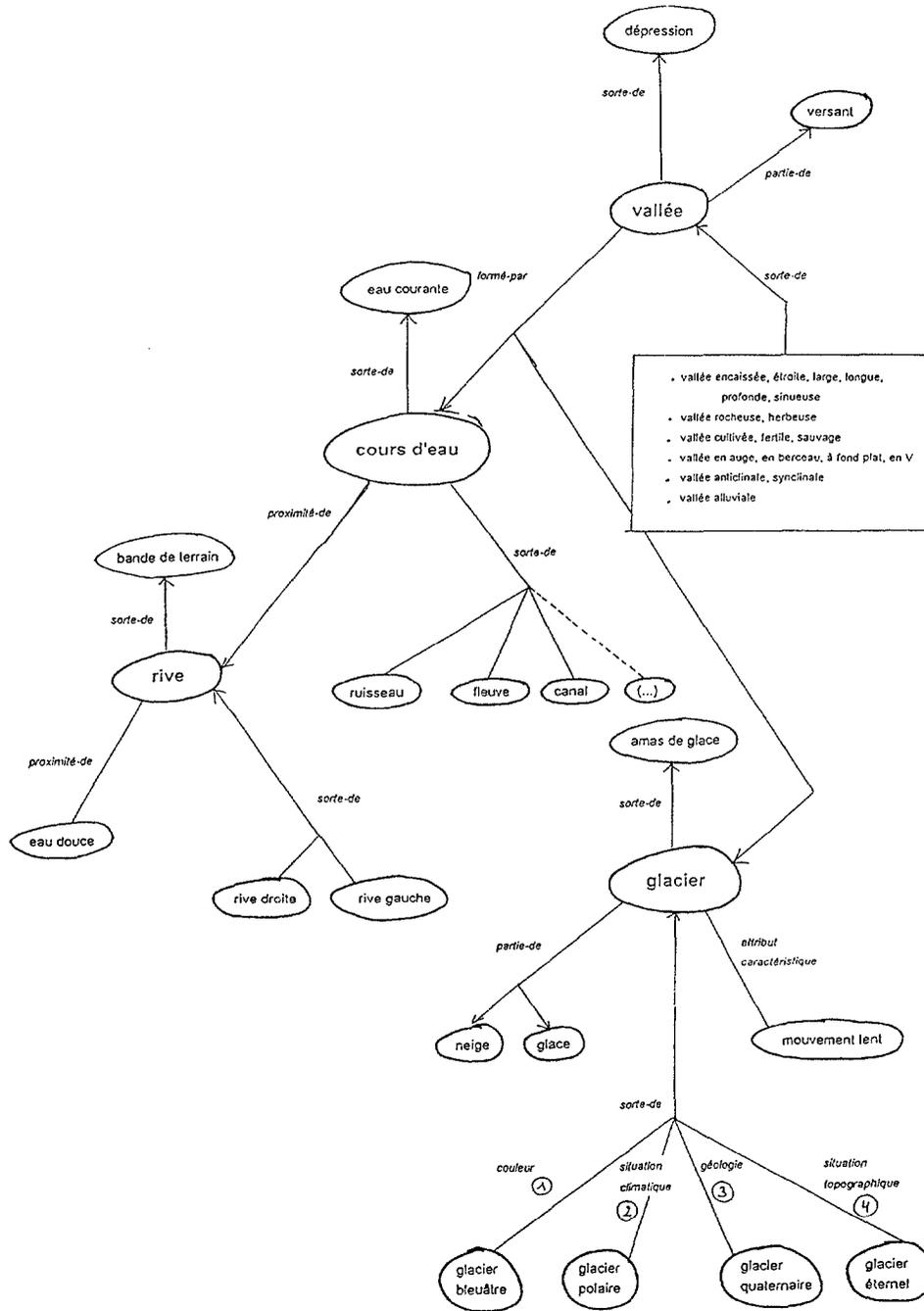


Figure 4:
RST de vallée et glacier à partir du TLF

3. LA NOTION DE BCT

La notion de base de connaissances terminologiques (BCT) est un prolongement naturel de la notion de base de connaissances et de système à base de connaissances (SBC). On attribue généralement la paternité de la notion de BCT à I. Meyer. Elle définit la BCT (en anglais TKB pour *terminological knowledge base*) comme une forme hybride à mi-chemin entre la base de données terminologiques (*termbank*) et la base de connaissances (*knowledge base*). La BCT contient toutes les informations d'une banque de données terminologiques classique augmentée pour chaque terme d'une forte dose d'informations conceptuelles hautement structurées (*a large and highly structured amount of conceptual information for each term*) (Meyer *et al.* 1991). Une BCT doit donc comporter une description explicite et structurée des concepts d'un domaine. Le passage de la base de données terminologiques (BDT) à la BCT oblige la terminologie à s'intéresser aux travaux effectués en IA et en TALN. À terme, toute BCT a pour vocation d'être intégrée à un SBC. Il en résulte qu'une BCT, même si elle est construite en dehors de tout objectif d'application spécifique, ne justifiera sa raison d'être que dans le cadre d'une application (construction d'une SBC par un cogniticien, par exemple). Précisons, comme le souligne D. Bourigault, que le cogniticien ne trouvera pas dans une BCT toute l'information conceptuelle dont il aura besoin pour construire un SBC particulier (Bourigault et Condamines 1995). En effet, il n'y trouvera aucun élément concernant la modélisation du raisonnement expert dans le domaine concerné et seulement certains éléments du réseau conceptuel terminologique seront pertinents pour le cogniticien vis-à-vis de l'application visée.

4. RÉALISATIONS

En 1995, l'idée d'appliquer les principes des réseaux sémantiques pour construire des bases de connaissances terminologiques n'est plus une idée entièrement originale. Un certain nombre de réalisations ont déjà vu le jour, d'autres projets sont en cours, et nous sommes actuellement partie prenante dans la réflexion et le développement d'au moins deux de ces projets.

4.1. COGNITERM

Le projet COGNITERM est sur l'établi depuis la fin des années 1980 au Laboratoire d'Intelligence Artificielle de l'Université d'Ottawa, sous la direction de Douglas Skuce. COGNITERM est un prototype de BCT construit à partir de l'outil informatique CODE (pour *Conceptually Oriented Descriptive Environment* ou système global de gestion de la connaissance). Dans cette BCT, les concepts sont représentés sous la forme de structures appelées *concept descriptors (CD)*. La visualisation graphique se présente sous la forme d'arbres et de réseaux de relations (hiérarchiques et non hiérarchiques). COGNITERM applique largement la notion de multidimensionnalité ; c'est pourquoi son interface graphique est dotée de fonctions de masquage (*masking capability*), qui permettent d'afficher uniquement les points de vue sélectionnés, et de fonctions de navigation qui permettent d'accéder sélectivement à certains types d'informations (termes, caractéristiques, etc.).

Le système est également doté d'une boîte à outils d'aide à la conception d'une BCT. Elle comprend en particulier une «matrice de comparaison de caractéristiques» (*Characteristic Comparison Matrix*) qui permet de construire par comparaison des définitions intentionnelles et de parfaire les classifications, ainsi qu'un outil de repérage d'incohérences fondé sur des mécanismes d'héritage (pointage de synonymes et quasi synonymes). CODE est également doté d'un outil de contrôle de la qualité qui permet de repérer différents niveaux de finalisation des entrées terminologiques. Ce dernier module est

particulièrement utile dans le cas où un grand nombre de terminologues sont mis à contribution pour alimenter la base et où ces contributions sont parallèlement soumises à un collège d'experts. Deux aspects problématiques de toutes les terminologies sont spécifiquement pris en compte au cours des diverses phases de contrôle de la qualité : l'incohérence (*inconsistency*), phénomène qui se manifeste par l'observation chez des locuteurs différents de définitions distinctes et incompatibles, et la « surcharge terminologique » (*overloading*) qui résulte de l'utilisation chez le même locuteur ou groupe de locuteurs de définitions multiples pour le même terme. En amont, CODE aide le terminologue cognicien à construire des structures systématiques des domaines terminologiques à représenter. L'interface graphique permet au terminologue d'examiner différentes possibilités de classification, CODE signalant automatiquement toute présence multiple du même concept ou de la même caractéristique. Les concepteurs ont été particulièrement attentifs à la notion de multidimensionnalité (Bowker et Meyer 1993), à ses modes d'enregistrement et de présentation. CODE comporte par ailleurs un outil « d'assistance à la désignation des propriétés » (*property naming assistance*) qui affiche à la demande l'ensemble des désignations de propriétés employées. L'utilisateur peut afficher le texte complet d'une relation pour éclaircir son sens (« *to clarify what it means* » (Skuce 1991)). Ce que le système perd en clarté sur ce plan, il le gagne probablement en richesse informative. COGNITERM est un produit destiné aussi bien aux professionnels de la langue (traducteurs, rédacteurs techniques...) qu'aux non-professionnels (étudiants cherchant à se spécialiser dans un domaine de connaissances, gestionnaires de systèmes d'information...), c'est-à-dire aussi bien à des experts qu'à des non-experts d'un domaine, pour leur permettre de produire plus facilement des textes de spécialité. I. Meyer et D. Skuce s'appuient pour cela sur une collaboration tripartite : terminologues, spécialistes du domaine et cogniciens. Partant du constat qu'« aucune étude n'a cherché à déterminer avec exactitude ce que les différents types d'utilisateurs veulent savoir lorsqu'ils interrogent une banque de terminologie », I. Meyer prend l'hypothèse que « moins l'utilisateur s'y connaît dans un domaine, plus il comptera sur la banque de terminologie pour l'éclairer » (Meyer et McHaffie 1993). Cet éclairage nécessaire à la compréhension approfondie de notions techniques ne peut provenir des seules fiches terminologiques d'une banque ou des réseaux d'une BCT. Le recours (ou le retour) aux corpus initiaux devient un prolongement utile sinon indispensable. Il est aujourd'hui relativement aisé grâce à la constitution de corpus électroniques de textes de spécialité et de fonctions de navigation de type hypertexte.

4.2. Projet Matra Marconi Space

Ce projet conjoint entre la société Matra Marconi Space et l'équipe de Recherches en Syntaxe et en Sémantique de l'Université de Toulouse-Le Mirail a abouti en 1994 à la constitution d'une base de connaissances terminologiques dans le domaine de l'industrie aéronautique et spatiale¹. Cette base de connaissances est destinée à équiper un poste d'aide à la rédaction de documents techniques dont la réalisation est assurée par Matra Marconi Space. Le modèle de représentation retenu comporte à la fois des éléments terminologiques et lexicologiques. Les premiers permettent la construction d'un réseau conceptuel (en particulier à partir des relations « *est-un* » et « *partie-de* » avec les liens *est générique de*, *est spécifique de*, *est composé de* et *fait partie de*), les seconds indiquent les conditions d'usage des termes en question (niveau de langue, type de document concerné, sous-domaine d'emploi, statut du terme (standard, terme-client, terme préconisé ou à éviter), en particulier avec le lien *est exclu de*). Ce modèle, parce qu'il est tributaire d'une application industrielle et d'un environnement d'utilisateurs clairement identifié, prend en compte des éléments qui ne relèvent pas strictement de la terminologie mais de la lexicologie, de la morphologie, de la syntaxe, de la pragmatique et du discours scientifique et technique

(Bras *et al.* 1992 ; Condamines 1993). Ce n'est certes pas un défaut mais indéniablement une source de complications. Ce dernier point peut expliquer la relative pauvreté actuelle de la partie conceptuelle et proprement sémantique du modèle (deux relations seulement). Les auteurs de ce système visent la conception d'un système d'assistance destiné aux rédacteurs techniques doté d'outils de vérification terminologique, de contrôle de cohérence et de détection d'ambiguïtés dans les textes techniques que ces derniers sont amenés à produire à l'intérieur de leur entreprise. La barre est placée très haut. Il n'en demeure pas moins, comme l'ont précisé ses concepteurs à l'occasion de la remise du prix CNRS-ANVIE, que ce projet est la preuve tangible que «des entreprises prennent conscience des surcoûts et des risques d'erreur que peuvent entraîner les problèmes de communication ou de mauvaise compréhension des langages spécialisés. Il témoigne de l'utilité de la linguistique pour aider les entreprises à résoudre les problèmes de maîtrise et de traitement de l'information scientifique et technique, notamment dans les secteurs de la haute technologie». L'enseignement qu'en tire le modeste modélisateur que nous représentons est qu'il est essentiel de recourir à des applications grandeur nature et *in situ* de ce type pour affiner des prototypes élaborés en laboratoire qui restent souvent trop éloignés des réalités de terrain.

5. PROJETS EN COURS

5.1. Projet de traitement terminologique du TLF

L'intérêt grandissant des communautés terminologique, lexicologique et cognitive pour les travaux ayant trait au développement de bases de connaissances nous a permis d'être personnellement partie prenante dans deux projets en cours. Le premier, que nous appellerons ICE (INaLF-CRIN-ERSS, c'est-à-dire Institut National à la Langue Française, Centre de Recherche en Informatique de Nancy et Équipe de Recherches en Syntaxe et Sémantique de Toulouse), et qui associe linguistes et informaticiens travaillant dans le domaine du traitement de la langue naturelle et de la représentation des connaissances, vise un double objectif : développer des méthodes et des outils informatiques pour aider les linguistes dans leurs travaux terminologiques et lexicographiques d'une part, rendre disponibles de nouvelles sources d'information pour les informaticiens qui travaillent dans le domaine de l'IA, d'autre part. L'équipe chargée de ce projet est partie d'un double constat de faiblesse : la base textuelle FRANTEXT et le TLF constituent des ressources trop peu exploitées par les informaticiens, bien qu'elles contiennent des données intéressantes pour la construction de BCT d'une part, les systèmes de TLN pèchent souvent par manque de connaissances du domaine qu'ils traitent d'autre part. Dans un premier temps, le projet consiste à proposer à l'INaLF un outil de constitution de BCT dans le cadre de la réalisation de dictionnaires de spécialité et de l'informatisation du TLF (*Trésor de la Langue Française*). Plus concrètement, il s'agit de mener de front un certain nombre de traitements terminologiques sur le TLF et sur un corpus de textes pour caractériser l'apport potentiel respectif du TLF et des textes de spécialité dans la réalisation d'une BCT. Ces traitements consistent à :

- sélectionner un domaine de spécialité (le domaine retenu est celui de la *géomorphologie*) ;
- repérer les termes de ce domaine dans le TLF et en extraire les définitions ;
- étiqueter les parties du discours dans les définitions ;
- déterminer les relations sémantiques entre ces termes et constituer une première base de connaissances.

Termes repérés et relations entre termes seront alors introduits dans un modèle formel de représentation des connaissances. L'intérêt d'un tel modèle dans le cadre de ce projet est

de donner aux utilisateurs un moyen de «raisonner» sur le contenu de la base en associant aux relations des propriétés logiques, et d'interroger la base sur l'interdépendance des termes dans le dictionnaire. C'est ici que nous envisageons de recourir aux RST ou à un modèle similaire. Les étapes énumérées ci-dessus seront appliquées conjointement à un corpus de textes du même domaine de spécialité (*géomorphologie*) afin de permettre la comparaison entre les deux BCT réalisées à partir de l'un ou de l'autre support. L'analyse comparative portera sur les différences de nomenclature, sur les variations entre termes sur le plan des définitions et des relations. Le modèle de représentation envisagé (qui n'est pas encore arrêté au moment de la rédaction de ces lignes) est tributaire de la capacité expressive de la BCT et des requêtes que l'on souhaite effectuer sur celle-ci. Compte tenu des applications envisagées, la validation de la BCT obtenue se fera selon trois axes complémentaires : complétude, cohérence et «sphère de pertinence». Le test de complétude vise à vérifier qu'aucune notion essentielle n'a été oubliée et que tous les sous-réseaux de la base sont interconnectés. Le test de cohérence permet de vérifier que la BCT ne contient pas une information et son contraire. La notion de sphère de pertinence (Toussaint 1992) peut être caractérisée par la recherche d'une réponse valide à la question «Comment s'y prendre pour ne pas extraire toute la BCT lorsque l'on s'intéresse aux relations qu'entretient un concept donné avec un nombre limité d'autres concepts?». Le terminologue sait intuitivement qu'à partir d'une certaine «distance» exprimée en nombre de relations parcourues, les concepts deviennent de moins en moins significatifs par rapport au concept initial. La notion de sphère de pertinence vise en fait à optimiser les recherches dans les BCT. L'intérêt d'un tel projet réside principalement dans le fait qu'il va largement au-delà de la problématique du modèle de représentation pour s'insérer dans une situation concrète de réalisation de BCT qui fait appel à une gamme très large de techniques issues de l'ingénierie linguistique.

5.2. Terminologie et IA : groupe PRC-IA

Le groupe «Terminologie et IA» du PRC-IA, créé en 1993 et avalisé en 1994, s'est fixé comme axe de travail principal l'établissement d'une collaboration entre chercheurs en linguistique et en IA par le biais de la terminologie². La convergence de ces deux disciplines se focalise dans ce cadre autour de la notion de base de connaissances : en partant des SBC (systèmes à base de connaissances) de l'IA et des BDT de la terminologie, la rencontre entre chercheurs en IA et linguistes terminologues se situe prioritairement autour de la définition du concept de base de connaissances terminologiques (BCT). La communauté des terminologues est de plus en plus largement convaincue de la nécessité d'intégrer dans ses banques de terminologie des données linguistiques et conceptuelles sur la connaissance véhiculée par les termes en tant que désignateurs de concepts. Les BCT qui en résulteront constitueront des gisements inestimables pour les cognitivistes dont la tâche est de construire des SBC (Ozman 1992b). Les activités actuelles du groupe s'articulent autour de quatre axes de réflexion principaux :

- **des problèmes théoriques** : quelles positions adopter pour développer une collaboration efficace entre linguistes et chercheurs en IA ;
- **BCT et SBC** : les deux notions sont à distinguer, surtout si l'on considère que la première est destinée à servir de support à la réalisation de la seconde. Le travail de réflexion du groupe consiste sur ce point à définir la structure et les spécifications requises d'une BCT pour qu'elle puisse être exploitée dans le processus de construction d'un SBC ;
- **BCT et système de TLN** : les systèmes de TLN ont besoin de «connaissances terminologiques» au moment de l'analyse syntaxique (reconnaissance des UT) et de l'analyse sémantique (connaissances sémantiques véhiculées par les termes) pour interpréter les phrases et les textes qui leur sont soumis ;

- **BCT et formalismes de représentation des connaissances** : à partir d'une étude menée sur un certain nombre de formalismes existants (graphes conceptuels, réseaux de type KL-ONE...), le groupe envisage de proposer un formalisme de représentation original compatible avec les nécessités d'une BCT et les contraintes d'un SBC.

L'objectif actuel du groupe consiste à concrétiser ces points de réflexion autour de la réalisation d'un prototype de BCT dans le domaine médical (sous-domaine des maladies coronariennes).

CONCLUSION : LES LIMITES DU FORMALISME

Ces premiers travaux ont permis de mettre en évidence un certain nombre de limites inhérentes à ce formalisme de représentation. Notre modèle ne prend pas en compte, ou insuffisamment, la négation, les modalités discursives et autres formes de restriction exprimées dans le discours scientifique et répercutées au niveau des définitions dans les terminologies. Prenons à titre d'exemple la définition suivante de *ordinateur de cinquième génération* extraite de notre corpus de référence :

Machine de traitement logique ou symbolique très performante en termes d'interaction homme-machine destinée en particulier au traitement des connaissances. (Otman 1991)

Cette définition comporte au moins trois éléments exprimant une restriction : la conjonction «ou» (logique **ou** symbolique), l'adverbe «très» (**très** performante) et la locution adverbiale «en particulier». De plus, notre modèle s'appuie trop largement sur une terminologie hors discours et ne prend pas en compte la dimension phraséologique de la discipline terminologique et, en particulier, l'existence de verbes supports obligatoires ou privilégiés pour certains termes :

exemples : **attribuer** une valeur par défaut
 adresser un message
 connecter un processeur
 tracer un graphe
 quantifier une variable...

Nous pensons ici tout particulièrement aux notions de classes d'objets et d'opérateurs verbaux développés par Gross (1994). Il est indéniable que si une BCT permettait d'accéder à la liste de tous les «objets» d'un domaine terminologique qui acceptent tel verbe en construction (tracer : graphe, arbre, réseau...) et, à l'inverse, de tous les verbes qui peuvent entrer en combinaison avec tel terme (graphe : tracer, parcourir, remonter dans...), elle s'en trouverait énormément enrichie aux yeux de certains utilisateurs spécialisés. Dans les exemples cités, la démarche a consisté à vérifier la validité du modèle et la qualité terminologique des corpus soumis à l'analyse. C'est une étape prioritaire nécessaire à la réalisation d'un modèle et à la mise en place d'un outil informatique d'aide à la conception terminologique. La seconde étape, que nous qualifierons d'opérationnelle, consiste à proposer cet outil à des terminologues afin qu'ils s'en servent au moment de la conception même de leurs bases terminologiques, en particulier dans les phases de :

- rédaction de définitions ;
- recherche d'hyperonymes, d'hyponymes et de co-hyponymes ;
- traitement de la synonymie (réelle ou apparente) ;
- traitement de la cohérence entre co-hyponymes ;
- établissement de listes systématiques ;
- élimination de redondances, d'incohérences et de termes non liés au domaine.

La terminologie ne concerne pas seulement les terminologues. Elle concerne — et préoccupe — de plus en plus traducteurs, rédacteurs techniques, documentalistes, indexeurs, mais également des professionnels comme les cognitivistes et les administrateurs de systèmes d'information. C'est pourquoi notre objectif est de travailler en collaboration avec ces experts utilisateurs de terminologies afin qu'à l'aube du siècle prochain, des BCT soient intégrées à des systèmes à bases de connaissances et des stations de travail de communicateurs techniques.

Notes

- * Cet article est issu d'une communication présentée par l'auteur aux IV^{es} Journées scientifiques du réseau «Lexicologie, terminologie, traduction» de l'AUPELF-UREF (Lyon, France, 28, 29, 30 septembre 1995).
1. Ce projet a valu à ses auteurs le prix CNRS et ANVIE 1994 de la valorisation.
 2. Le groupe se compose à l'heure actuelle des personnes suivantes : Didier Bourigault (EDF-Clamart), Anne Condamines (ERSS-Toulouse), Brigitte Biebow et Sylvie Szulman (LIPN-Paris-Nord), Benoît Habert (ENS Fontenay Saint-Cloud), Chantal Enguehard (IRIN-Nantes), François Rousselot (ERIC-USHS Strasbourg), Jean Royauté (INIST-Nancy), Yannick Toussaint (CRIN-Nancy), Pierre Zweigenbaum, Bruno Bachimont et Jean Charlet (INSERM-Paris) et Gabriel OTMAN (LLI-CNRS-Paris-Nord).

RÉFÉRENCES

- BOURIGAUT, D. et A. CONDAMINES (1995) : «Réflexions sur le concept de base de connaissances terminologique», *Actes des 5^{es} journées nationales*, PRC-GDR IA, TEKNEA, pp. 425-443.
- BOWKER, L. et I. MEYER (1993) : «"Textbook" Concept System: Handling Multidimensionality in a New Generation of Term Banks», *Actes du colloque TKE 93*, Francfort, Indeks Verlag, pp. 123-137.
- BRAS, M. *et al.* (1992) : «Towards an Electronic Dictionary of the Space Technical Domain», document interne, International Astronautical Federation.
- CONDAMINES, A. (1993) : «Terminology Between Language and Knowledge: an Example of Terminological Knowledge Base», *Actes du colloque TKE 93*, Francfort, Indeks Verlag, pp. 316-323.
- GROSS, G. (1994) : «Classes d'objets et description des verbes», *Langages*, n° 115, septembre 1994, Paris, Larousse, pp. 15-30.
- HUMBLEY, J. (1994) : «L'exploitation des bases de données terminologiques pour la lexicographie spécialisée», D. Candel (dir.), *Français scientifique et technique et dictionnaire de langue*, Paris, INaLF, Didier Érudition, pp. 37-51.
- LERAT, P. (1995) : *Les langues spécialisées*, coll. «Linguistique nouvelle», Paris, PUF, 206 p.
- MEYER, I. (1995) : «Constructing a Knowledge-based Term Bank: Fundamentals and Implications», *Actes du colloque Terminology and Documentation in Specialized Communication*, Hull.
- MEYER, I. et B. McHAFFIE (1993) : «De la focalisation à l'amplification : nouvelles perspectives de représentation des données terminologiques», A. Clas et P. Bouillon (dir.), *TA-TAO : recherches de pointe et applications immédiates*, Actes des Troisièmes Journées scientifiques, Montréal du 30 septembre au 2 octobre 1993, Beyrouth, AUPELF-UREF et FMA, pp. 425-440.
- OTMAN, G. (1990) : «Semantic Network in Terminology: the Case of AI», H. Czap (Ed.), *TKE'90, Terminology and Language Engineering*, vol. 1, Francfort, Indeks Verlag, pp. 210-221.
- OTMAN, G. (1991) : *Vocabulaire de l'intelligence artificielle*, Nanterre, Éditions EC2.
- OTMAN, G. (1992a) : «Des réseaux sémantico-terminologiques pour des banques de données terminologiques intelligentes», *Actes du XV^e Congrès international des linguistes*, Québec, Université Laval.
- OTMAN, G. (1992b) : «Knowledge Engineers, do not Neglect Terminology — Collection and Formalization of Expert Terminological Information by the Knowledge Engineer», A. M. Tjoa et I. Ramos (Eds), *Actes du colloque Database and Expert Systems Applications*, Vienne, Springer-Verlag, pp. 409-414.
- OTMAN, G. (1992c) : «Une approche résolument sémantique de la terminologie», *Actes du 3^e symposium d'Infoterm : Travail terminologique dans les domaines de spécialités*, Vienne, TermNet, pp. 536-554.
- OTMAN, G. (1995) : *Les représentations sémantiques en terminologie*, Thèse de doctorat, Paris IV, Sorbonne, 357 p.
- PARADIS, C. et P. AUGER (1987) : «La terminotique ou la terminologie à l'ère de l'informatique», *Meta*, 32 (2), pp. 102-110.
- QUILLIAN, R. (1968) : «Semantic Memory», M. Minsky (Ed.), *Semantic Information Processing*, Cambridge (Mass.), MIT Press, pp. 227-270.
- TOUSSAINT, Y. (1992) : *Méthodes informatiques et linguistiques pour l'aide à la spécification de logiciel*, Thèse de doctorat, Université Paul-Sabatier, Toulouse, octobre 1992.