



Terminologie et représentation des connaissances

Anne CONDAMINES

Équipe de Recherches en Syntaxe et Sémantique
(ERSS, URA 1033)

Maison de la Recherche, Université Toulouse le Mirail
5, allées Antonio Machado
31058 Toulouse cedex, France.

Ce dossier était en cours de rédaction quand j'ai rencontré dans un colloque sur les terminologies, où nous étions tous deux conférenciers invités, Anne Condamines¹. Son exposé était doublement intéressant. D'une part, il portait sur les difficultés de communication et de traduction (entre professionnels du secteur de l'industrie aéronautique et spatiale) qu'entraîne la polysémie de certains concepts techniques. D'autre part, il était accompagné de transparents qui ressemblaient fortement à des cartes conceptuelles. Comment et pourquoi une linguiste est-elle amenée à spatialiser des concepts dûment étiquetés et à les relier par des ponts ? De plus, il est rare que des linguistes centrent leurs recherches sur les langues de spécialité et plus particulièrement les terminologies techniques mobilisées par les chercheurs, ingénieurs et techniciens de l'industrie. Cet éclairage était évidemment stimulant pour une revue qui a pour projet de s'intéresser, entre autres, à la communication de ces savoirs techniques. Mais surtout il nous a paru utile de compléter le point de vue de l'intelligence artificielle par celui d'une linguiste qui, il est vrai, est par ailleurs soucieuse d'intégrer certaines des méthodes des informaticiens.

D. Jacobi

1. Anne Condamines et l'équipe CNRS à laquelle elle appartient ont reçu pour leurs travaux le prix de la valorisation 1994 du département SHS du CNRS et de l'ANVIE.

Résumé

La terminologie est à un moment crucial de son histoire. La rencontre avec l'intelligence artificielle, les besoins accrus des entreprises en ce qui concerne la communication technique, devraient l'aider à structurer et organiser les recherches futures dans ce domaine. Un rapprochement avec la linguistique pourrait être fructueux, et permettrait de trouver un terrain d'entente avec l'intelligence artificielle. La terminologie et l'intelligence artificielle ont en commun de travailler sur la représentation des connaissances, la première via l'étude des concepts désignés par les termes, la seconde via l'étude de la formalisation de la connaissance. Cet article se propose de montrer comment et pourquoi la collaboration est possible entre ces deux disciplines.

Mots clés : *intelligence artificielle, représentation de la connaissance, réseau conceptuel, réseau sémantique, terminologie.*

Abstract

Terminology has reached an important moment in its development as a discipline. The confluence with artificial intelligence on the one hand, the increasing industrial needs for communication on the other could help terminology to structure its future researches. They could find a fruitful terrain in closer contacts with linguistics and, consequently, with artificial intelligence. Terminology and artificial intelligence share the same aim, that of knowledge representation : terminology via the study of concepts denominated by terms, and artificial intelligence via knowledge formalisation. This article proposes to show how and why the collaboration is possible between these two disciplines.

Key words : *artificial intelligence, conceptual networks, semantic networks, knowledge representation, terminology.*

Resumen

La terminología está en el momento crucial de su historia. El encuentro con la inteligencia artificial de una parte, y, por otra parte las necesidades del incremento industrial en lo que concierne a la comunicación técnica, debería ayudarla a estructurar y organizar las futuras investigaciones en este dominio. Ellas deberían encontrar un terreno fértil en un acercamiento con la lingüística, acercamiento que les permitirá encontrar un terreno de entendimiento con la inteligencia artificial. La terminología y la inteligencia artificial tienen en común el trabajar sobre la representación del conocimiento, la primera, via el estudio de los conceptos designados por los términos, la segunda, via el estudio de la formalización del conocimiento. El artículo se propone mostrar cómo y por qué la colaboración es posible entre estas dos disciplinas.

Palabras claves : *inteligencia artificial, representación del conocimiento, red conceptual, red semántico, terminología.*

INTRODUCTION

La rencontre entre la terminologie et la représentation des connaissances, issues de courants de recherches différents et poursuivant apparemment des objectifs différents, ne peut se faire sans des conditions particulièrement favorables. Il semble que ces conditions se mettent en place et que ceci soit à rattacher, en grande partie, à l'émergence d'un besoin en entreprise. Dans ce type de lieu cohabitent, en s'ignorant le plus souvent, des services de traduction, où se pratique la terminologie, et des départements de recherche et développement en informatique linguistique et documentaire et en intelligence artificielle, qui, tout comme les services de traduction, sont amenés à traiter des textes spécialisés ; dans les deux cas, les travaux visent une amélioration de la communication technique.

Dans le même temps, une collaboration s'est développée dans les laboratoires de recherche entre linguistes et informaticiens, autour de la problématique du traitement automatique du langage naturel. Cette collaboration commence à porter ses fruits et permet de proposer un début de réponse aux entreprises en permettant une mise en contact des travaux sur la terminologie et sur l'informatique.

L'état des recherches sur la terminologie et l'intelligence artificielle sera présenté dans cet article, ainsi que les nécessaires évolutions auxquelles devraient conduire la rencontre terminologie/intelligence artificielle. Puis nous exposerons un modèle de Base de Connaissance Terminologique, manifestant la rencontre entre ces deux disciplines.

1. TERMINOLOGIE, REPRÉSENTATION DES CONNAISSANCES : LES POSSIBILITÉS DE RENCONTRE

Héritière des travaux sur la classification du siècle dernier, la terminologie s'est constituée en discipline autonome grâce à Wuster, il y a une cinquantaine d'années. Depuis, l'activité de terminologie s'est essentiellement développée dans des objectifs de traduction ou de normalisation non dénués d'ambitions politiques (protection des langues) ou économiques (meilleur rendement des traducteurs et éventuellement, tentative de faire imposer les termes en cours dans une entreprise par exemple). Ces développements se sont souvent faits avec une vision un peu naïve de la terminologie qui ne serait qu'une sorte de représentation parfaite d'un système conceptuel sous-jacent :

“[Pour les philosophes des sciences], la connaissance scientifique procédant du raisonnement logique, il est possible de bâtir un système sémiotique optimal entièrement fondé sur la logique. L'unité minimale est le terme, “pur” de toute connotation, univoque, monoréférentiel et précis.” (Slodzian, 1994)

“Si le positivisme logique est aujourd'hui largement rejeté, certaines de ses conclusions continuent de prévaloir localement. La terminologie tra-

duit dans ses préceptes les plus dogmatiques la survivance d'un positivisme révolu." (ibidem)

Le flou le plus regrettable dans les travaux existants vient certainement de ce que l'objet de la terminologie n'est pas suffisamment clairement défini ; en particulier, les réflexions autour de la définition même du terme ne nous semblent pas assez poussées. Les possibilités offertes par l'informatique mais aussi les nouvelles questions qu'elle pose ne font que renforcer ce malaise.

Aujourd'hui, la terminologie est certainement à une étape importante de son histoire et nous pensons que la confrontation avec d'autres disciplines lui permettra d'élargir et de mieux comprendre son champ d'investigation.

Pour nous, une des disciplines avec laquelle la confrontation est la plus fructueuse, c'est l'informatique et, plus particulièrement, l'intelligence artificielle. De cette confrontation, des possibilités et des limites qu'elle montre, il semble qu'on peut d'ores et déjà tirer l'idée que, tant pour son avenir propre que pour pouvoir s'ouvrir à la fois à de nouvelles disciplines et à de nouvelles applications, la terminologie doit s'ancrer encore davantage dans une science qui lui permettra d'aller plus loin dans l'explication des fonctionnements : la linguistique.

1.1. Le nécessaire ancrage de la terminologie dans la linguistique

Deux éléments, présentés ci-dessous, semblent avoir éloigné les travaux sur la terminologie de la linguistique, et ce malgré le rapprochement préconisé par Wuster.

1.1.1. Héritant du positivisme du XIX^e siècle, les terminologues ont parfois tendance à ne considérer les termes que comme des étiquettes d'éléments de la réalité :

"Par l'intermédiaire du sens, le terme est en rapport avec le référent qu'il représente et dont on parle surtout s'il s'agit d'objets matériels." (Kocourek, 1982)

Seule la fonction de représentation du terme est alors considérée ; l'étiquetage du référent est fixé hors discours et de façon permanente. Une telle approche suppose que le terme-étiquette peut apparaître dans n'importe quel contexte sans que son "sens" en soit affecté. En réalité, les points de vue sur les référents entraînent des conceptualisations différentes qui peuvent amener à des dénominations différentes.

Ainsi, dans la terminologie du spatial, le satellite artificiel qui correspond à un référent unique, peut être vu comme objet dont il faut simuler le comportement, objet qu'il faut réellement envoyer dans l'espace, objet qu'il faut maintenir, etc. Suivant les départements concernés, le point de vue n'est pas le même et les contextes d'apparition seront aussi très différents. Même si, dans ce cas, il n'y a pas de variations dans la dénomination, il faut bien donner un sens au fait que les contextes d'apparition sont très différents. Nous considérons qu'il y a alors plusieurs concepts en jeu.

Par ailleurs, toujours chez Kocourek, il est étonnant de voir associer le “sens” d’un terme qui renvoie à un fonctionnement linguistique au “réfèrent” que, depuis Saussure, on considère comme n’appartenant pas au champ d’investigation de la linguistique. N’est-ce-pas effectuer un sérieux raccourci que d’établir ainsi un lien direct entre terme et réfèrent ?

Nous pensons que, s’intégrant dans le système linguistique d’une langue de la même façon que n’importe quel mot¹, le terme doit être avant tout considéré comme un signe linguistique.

Il est vrai cependant que, parce que la terminologie est toujours associée à un domaine particulier, il semble possible de considérer que la plupart des termes entretiennent une relation stable avec les concepts (et non pas les référents) du domaine ; on peut ainsi faire l’hypothèse que les termes peuvent constituer des clés d’accès à la connaissance du domaine. *Via* l’examen linguistique des termes et de leurs contextes d’apparition, on peut accéder aux éléments conceptuels stabilisés du domaine :

“Bien entendu, en règle générale, un concept donné a pour pendant, dans une langue donnée, un signifié qui en tient lieu dans les limites de cette langue, mais la pensée n’est pas la langue et réciproquement.”
(Lerat, 1988)

Ne serait-ce que pour rendre l’analyse terminologique opérationnelle, nous pensons qu’il faut distinguer réfèrent, concept et signifié.

Le réfèrent concerne la perception et ne fait pas partie d’un système.

Le concept concerne la conceptualisation et fait partie d’un système.

Le signifié ne concerne que le linguistique et fait partie d’un système.

Le novice qui rencontre une partie d’un satellite, par exemple la charge utile, perçoit un objet, mais il ne sait pas à quoi il sert ni comment il s’intègre au système du satellite (il n’a accès qu’au réfèrent). Un expert, quelle que soit sa langue, sait parfaitement la fonction et la façon d’utiliser cet objet, il sait l’intégrer à un système de concepts (il en a une bonne conceptualisation). Ce même expert sait donner un nom à cet objet dans sa langue et, probablement, dans d’autres langues (il connaît le signe linguistique adéquat).

Il est clair que, tout comme il existe des relations entre signifié et concept, il en existe entre concept et réfèrent. Ainsi, il est entendu que notre façon de percevoir est influencée par notre façon de concevoir, elle-même étant influencée par notre façon de dire. Mais il convient de ne pas confondre les niveaux et de mesurer à quelles conditions on peut établir des équivalences entre eux.

1.1.2. Dans son entreprise pour définir la langue comme objet d’étude de la linguistique, Saussure a délibérément exclu tout ce qui renvoyait à l’énon-

1. Dans la majorité des cas, les termes s’insèrent en discours en obéissant à toutes les règles linguistiques, aussi bien en ce qui concerne l’accord que le fonctionnement syntaxique et sémantique (les seules exceptions concernent les sigles et les symboles chimiques, considérés comme termes, qui ne prennent pas les marques du féminin et du pluriel).

ciation et à la prise en compte du locuteur. Ce n'est que depuis quelques dizaines d'années que les travaux sur l'énonciation et la sociolinguistique se sont intéressés aux conditions de production de la langue, et on peut noter des différences dans ces productions, en fonction de la provenance sociale par exemple. Au niveau d'un domaine particulier, on peut faire l'hypothèse forte que c'est la différence de compétences qui sera manifestée dans les productions langagières ; on pourra distinguer les productions de novices des productions d'experts, en particulier par le fait que ces derniers utiliseront beaucoup plus de termes spécifiques.

Pourtant, la terminologie, directement concernée par la définition d'un domaine de compétences (et donc, d'un domaine de production langagière) a très peu bénéficié des apports de ces recherches¹. L'explication tient peut-être à la difficulté des linguistes d'aller sur le terrain même des productions terminologiques, c'est-à-dire, pour l'essentiel de la terminologie technique, les entreprises². Dans les entreprises, lorsque des travaux de terminologie existent, ils sont, la plupart du temps, cantonnés dans les services de traduction et uniquement considérés comme une pratique visant à recueillir et traiter les termes au fur et à mesure des difficultés de traduction rencontrées.

Cependant, depuis quelques années, conscients des besoins nouveaux en terminologie, reliés notamment à des difficultés dans la communication technique ou à la volonté de mémoriser leur propre savoir-faire, certaines grandes entreprises ont mis en œuvre des projets d'envergure dans lesquels les ressources terminologiques constituent un des éléments majeurs. C'est le cas en particulier pour toutes les entreprises qui ont entrepris des projets de Traitement Automatique du Langage Naturel (TALN) ou de représentation des connaissances. Le plus souvent, ces entreprises participent à des projets internationaux visant à une définition commune des besoins et à la constitution d'outils (par exemple, dans le projet EUROLANG, d'outils d'aide à la traduction automatique), à charge pour chaque partenaire de recueillir sa terminologie et les informations nécessaires pour que l'outil fonctionne dans son entreprise. Pour n'en citer que quelques-unes, EDF, AÉROSPATIALE, SIEMENS participent à des projets de ce type. Dans un certain nombre de cas, les responsables de ces projets ont une formation en linguistique, ce qui est le signe d'une mutation puisque, jusqu'à présent, les terminologues avaient une formation de traducteurs.

La terminologie a tout à gagner à se rapprocher de la linguistique (tout en gardant ses spécificités). Tout d'abord, elle y gagnera du point de vue méthodologique ; la linguistique dispose de méthodes d'analyse qui trouvent tout à fait leur place dans la description de terminologie. Ainsi, l'analyse des contextes permet de mieux cerner un mot dont on ne connaît pas le sens, en

1. Il faut noter cependant qu'il existe un courant de réflexion sur la "socioterminologie", en particulier à l'université de Rouen (Gaudin, 1993).

2. La psychologie n'a pas hésité, elle, à pénétrer le tissu industriel depuis plusieurs dizaines d'années ce qui a conduit à la création d'une nouvelle discipline : l'ergonomie. Or, il n'est pas rare que, sur des problèmes qui relèvent nettement de l'étude de la langue, les entreprises fassent appel aux ergonomes, déjà sur place, mais qui n'ont peut-être pas toujours toutes les compétences qui seraient requises.

faisant appel aux connaissances linguistiques que l'on a sur le reste de la phrase ; elle permet aussi de repérer les relations sémantiques qu'un mot entretient avec d'autres mots du même paradigme (relation de hiérarchie essentiellement ; pour plus de précision, cf. Condamines, 1993).

Presque tout reste à faire pour adapter les méthodes de l'analyse linguistique à l'analyse des termes. Le fonctionnement linguistique des termes est mal connu. Pour ne prendre qu'un exemple, nombreux sont les termes qui sont aussi des mots de la langue courante ; quelles spécificités dans leur comportement en contexte font qu'on peut repérer un fonctionnement terminologique (qui permet d'établir un lien privilégié avec un concept), et pas seulement un fonctionnement linguistique ? Nous proposons par ailleurs (Condamines, à paraître) une première liste de critères qui permettent de repérer les termes. D'autre part, dans le rapprochement avec la linguistique, la terminologie sera gagnante parce qu'elle se trouvera sur un terrain d'entente avec d'autres disciplines et, en particulier, avec l'intelligence artificielle (cf. ci-dessous). Enfin, le rattachement à la linguistique ouvrira de nouveaux horizons à la terminologie et, du point de vue applicatif (aide à la rédaction de documents techniques, enseignement assisté par ordinateur par exemple), l'aidera à mieux adapter le travail aux besoins des entreprises.

1.2. Représentation des connaissances : TALN, IA

Presque tous les types d'utilisation de la terminologie dans des projets applicatifs font appel à des ressources qui doivent rendre compte du double fonctionnement du terme : comme signe linguistique et comme clé d'accès à une connaissance spécialisée. Dans ce paragraphe, nous souhaitons montrer les liens qui peuvent s'établir entre les travaux sur la terminologie et les travaux existant en informatique sur la représentation des connaissances. Deux types de courants de recherche, eux-mêmes fortement corrélés, sont engagés sur la représentation des connaissances : le Traitement Automatique du Langage Naturel (TALN) et l'Intelligence Artificielle (IA).

1.2.1. TALN

Depuis les débuts du traitement automatique du langage naturel, la plupart des chercheurs savent qu'ils ne peuvent (ne pourront) réussir dans leur entreprise que s'ils se focalisent sur des domaines de connaissance particuliers. Or, cette idée de domaine de connaissance est un des éléments qui permettent de caractériser la terminologie. Et en effet, la plupart des mots ou groupes de mots qui doivent être traités dans ces systèmes de TALN sont des termes ("capteur d'acquisition solaire", "étagiste", "bus fond de panier" sont des termes extraits de la terminologie de Matra Marconi Space (MMS), dont le recueil et le traitement sont indispensables dans un système de traitement automatique de la langue).

De plus, ces systèmes ont besoin des deux types de données que l'on retrouve, de façon plus ou moins développée, dans les Bases de Données Terminologiques (BDT) : des données linguistiques (nature, genre, types de

constructions admis...) et des données sur la connaissance du monde (essentiellement, des données sur les relations que les concepts désignés par les termes entretiennent entre eux).

Cependant, par rapport à la terminologie classique, l'intégration à un système de traitement automatique demande que soient posées un certain nombre de questions qui ne sont pas sans conséquence sur la façon d'envisager le traitement terminologique.

– Toutes les formes de réalisations linguistiques doivent être considérées, et pas seulement les formes nominales comme c'est souvent le cas dans les BDT existantes : ainsi, dans la terminologie de MMS, on trouve non seulement des noms, en grand nombre il est vrai ("banc d'intégration satellite"), mais aussi des formes verbales ("surveiller passivement"), des formes adjectivales ("différé en temps immédiat"), des formes adverbiales ("au fil de l'eau").

– Il faut s'interroger sérieusement sur les rapports entre le lexique général et la terminologie. D'un point de vue théorique, la principale question qui se pose concerne sans doute la notion de concept, assez facile à appréhender dans un domaine de compétence particulier, mais qui devient plus abstraite quand cohabitent compétence particulière et compétence générale, concepts experts et concepts généraux, termes et mots généraux.

– Enfin, une autre question devient incontournable, celle de l'établissement de relations entre termes, plus précisément, entre concepts dénommés par des termes. Dans le meilleur des cas, en terminologie, seules sont considérées les relations paradigmatiques de type *est-une-sortie-de* ou *est-une-partie-de*. Or, ces relations s'avèrent insuffisantes pour le TALN. Pour prendre le cas de la résolution automatique d'anaphore, certains cas nécessitent qu'aient été enregistrées de nombreuses données pas toujours faciles à définir *a priori*. Ainsi dans l'exemple :

"Paul s'est promené au Parc Borelli, il a vu une fleur magnifique, le jardinier lui a dit que c'était une amaryllis".

Pour résoudre l'anaphore concernant le défini "le", le système doit savoir le lien qui existe entre "fleur", "jardinier" et, probablement, "parc". On est loin des relations de hiérarchie, paradigmatiques, on est dans le syntagmatique où les relations sont beaucoup plus difficiles à analyser, à prévoir. Cependant, on peut penser que, dans un domaine de connaissance particulier, il est plus facile de définir ces relations (c'est d'ailleurs une des raisons pour lesquelles les systèmes de TALN ne fonctionnent "bien" que pour des langues de spécialité).

1.2.2. Intelligence artificielle

Depuis quelque temps, des voix se sont élevées pour réclamer une plus grande prise en compte des études de terminologie par les chercheurs en IA ; plus rarement, des chercheurs en IA ont encouragé leurs collègues à regarder du côté des résultats sur la terminologie (Ahmad, 1993). Le plus souvent, une commune indifférence règne, ce qui est regrettable. Dans le processus de construction de Systèmes à Base de Connaissance, propre à l'IA, on peut

distinguer deux étapes : l'acquisition de connaissance et la représentation des connaissances. À ces deux étapes-clés, la prise en compte de travaux sur la terminologie pourrait avoir un intérêt.

1.2.2.1. Acquisition de connaissance

Dans sa version actuelle, l'acquisition de connaissance a renoncé à passer directement du modèle cognitif de l'expert au système de représentation interne de la machine (Aussenac et al., 1992). Ce qui est en jeu actuellement, c'est de constituer un modèle de connaissance organisé en **modèle des données**, statique, et **modèle de raisonnement**, dynamique. Si le modèle de raisonnement peut être considéré comme étant propre à l'expert, le modèle des données doit être indépendant de l'expert et donc faire l'objet d'un accord à peu près consensuel dans la communauté experte du domaine concerné.

Or, n'est-ce pas justement ce consensus que recherchent les études sur la terminologie ? Dans sa démarche de définition des concepts manifestés par les termes, ces études ne cherchent-elles pas précisément à trouver les délimitations acceptables par l'ensemble des experts du domaine ? Et au bout du compte, la base de données terminologiques (ou plutôt une version évoluée de ce concept, cf. partie 2) ne contient-elle pas les éléments conceptuels stables à partir desquels peuvent s'élaborer des raisonnements individuels, qui, exprimés par des termes reconnus du domaine, sont compréhensibles par tout membre de la communauté ?

Bien que cela paraisse assez nettement, dans beaucoup de cas, il reste à en convaincre les chercheurs en IA.

1.2.2.2. Représentation des connaissances

Une bonne partie des systèmes de représentation de connaissances se situent dans la lignée initialisée par Quillian avec ses réseaux sémantiques¹.

"[Les réseaux sémantiques] sont formés sur la notion simple de graphe, formé de nœuds – représentant les concepts – reliés par des arcs. Cette unité de base qu'est le concept n'acquiert tout son sens que par les relations qui le lient aux autres concepts. Les arcs du graphe représentent alors des relations (généralement binaires) entre ces concepts. Les nœuds, comme les arcs sont étiquetés." (Sabah, 1988)

Dans la grande majorité des cas, parce que la connaissance représentée est spécifique d'un domaine, les étiquettes de ces nœuds correspondent à des termes et les relations retenues pourraient être considérées (du moins pour une part d'entre elles) dans la construction d'une base de données terminologiques.

Par ailleurs, il est regrettable que soit occulté, dans de tels systèmes de représentation, le fait que les termes utilisés pour étiqueter les nœuds et les

1. Rappelons que Quillian était un psycholinguiste qui voulait construire un modèle de mémoire sémantique.

relations sont aussi des signes linguistiques et ce, au moins pour deux raisons :

- au moment de la validation de ces réseaux, il faut que l'expert puisse retrouver la terminologie qu'il connaît pour se concentrer sur la seule évaluation du raisonnement,

- dans beaucoup de cas (sinon tous les cas), ces réseaux constituent les bases de connaissance qu'il faudrait associer à des systèmes de TALN tels que ceux dont nous parlions ci-dessus. Il est donc nécessaire que les deux types d'information, linguistiques (grammaticales) et cognitives (relations avec les autres concepts), puissent être attachées à chaque terme et que les termes soient intégrés à la base lexicale du système.

La rencontre entre terminologie et intelligence artificielle est possible car, nous espérons l'avoir montré, elle est bénéfique aux deux sciences, à condition que chacune accepte l'interdisciplinarité qui, bien souvent, avant d'être profitable est déstabilisante. L'intelligence artificielle peut trouver, dans la linguistique et la terminologie, des méthodes d'acquisition et de représentation de la connaissance. La terminologie peut trouver, dans l'intelligence artificielle, des modèles, voire des outils qui peuvent l'aider dans la constitution des bases de données terminologiques, essentiellement sans doute, pour construire des réseaux notionnels. Pour manifester cette convergence, le concept de Base de Connaissance Terminologique nous semble parfaitement approprié¹.

2. UN NOUVEAU CONCEPT : LA BASE DE CONNAISSANCE TERMINOLOGIQUE

Pour répondre aux nouveaux besoins concernant la terminologie, il est indispensable de créer de nouveaux modèles de stockage des données qui rendent compte à la fois du fonctionnement linguistique des termes et de leur lien fort avec les concepts du domaine². Ingrid Meyer, la première, (Meyer et al., 1992) a proposé le concept de Base de Connaissance Terminologique (BCT).

Nous concevons une Base de Connaissance Terminologique comme une amélioration d'une Base de Données Terminologiques classique enrichie par des relations conceptuelles. Ces relations concernent soit la partie descriptive du domaine (et l'on retrouve ici les relations bien connues est-un-type-de et a-pour-partie), soit la partie raisonnement, à condition qu'il emporte l'adhésion de l'ensemble des experts du domaine.

En effet, ce qui devrait être modélisé dans une BCT, c'est essentiellement ce qui est **consensuel** dans le domaine. Il est certain que puisque nous nous situons à la fois dans le domaine de la langue et celui de la connaissance, le consensus est très difficile à atteindre (un expert se distingue par l'origina-

1. L'auteur anime, avec Didier Bourigault, un groupe de réflexion sur le thème "Terminologie et Intelligence Artificielle" qui est financé par le Ministère de la Recherche.

2. Il est nécessaire aussi d'engager des études de fond sur le fonctionnement linguistique des termes et sur la façon dont ils établissent un lien avec la connaissance du domaine. Nous travaillons également sur ces problèmes mais nous ne les évoquerons pas ici.

lité de son raisonnement, non par un raisonnement partagé). Cependant, il semble raisonnable de penser que, dans tous les domaines, surtout s'il s'agit de domaines spécialisés, la possibilité d'une communication possible, *via* une langue spécialisée essentiellement caractérisée par des termes, est le signe du partage d'un minimum de connaissances communes. Plus précisément, le fait que la communication fonctionne, dans les domaines scientifiques, plutôt mieux que dans le domaine général, permet de faire l'hypothèse que les utilisateurs d'une langue de spécialité, à même niveau de compétence, peuvent s'entendre sur la définition des termes du domaine dans lequel ils évoluent.

C'est cette connaissance commune qui devrait être recueillie dans une Base de Connaissance Terminologique, cette BCT constituant un noyau de ressources qui pourrait être développé, adapté, enrichi en fonction des besoins.

2.1. Base de Connaissance Terminologique et Système à Base de Connaissance

Faisant l'hypothèse que la Base de Connaissance Terminologique constitue une ressource stable et fiable, il reste à montrer qu'en effet, la constitution d'un Système à Base de Connaissances (donc un système de stockage de données concernant le raisonnement d'un expert) est facilitée parce qu'il constitue une élaboration, un développement de la BCT. Si nous avons l'expérience de constitution de BCT (cf. 2.2), il faut préciser que leur utilisation pour la constitution de SBC n'a pas encore été testée.

Nous pouvons résumer ainsi les caractéristiques d'une Base de Connaissance Terminologique (BCT) et d'un Système à Base de Connaissance (SBC).

BCT	SBC
<ul style="list-style-type: none"> • Associée à un domaine • Rend compte de la connaissance consensuelle • Rend compte du fonctionnement linguistique et cognitif associé aux termes 	<ul style="list-style-type: none"> • Associée à un domaine, une application et un raisonnement • Rend compte de la connaissance statique et dynamique propre à un expert • Rend compte du fonctionnement cognitif

Sur la plupart des points, le SBC apparaît comme un enrichissement de la BCT, en particulier comme un affinement de la connaissance qui n'est plus générale mais spécifique à un expert. Cependant, la BCT présente l'avantage de permettre de conserver le lien avec la langue, ce qui est d'un grand intérêt (cf. partie 1). Il reste pourtant une question, fondamentale, à résoudre : les SBC sont des systèmes qui sont gérables par une machine, cela suppose que soient respectés un certain nombre de critères logiques (complétude, cohérence).

Or, il n'est pas certain que ces critères puissent être respectés au moment de la construction de la BCT. En particulier, la notion de complétude pose problème ; en effet, un domaine de connaissance n'est jamais isolable complètement de la connaissance générale ou d'autres connaissances spécifiques, elle prend ancrage dans la connaissance commune. En d'autres termes, comment peut-on envisager un réseau clos par lui-même ? Aux marges du domaine, il est très difficile de déterminer les concepts qui y appartiennent et ceux qui n'y appartiennent pas ; il est surtout impossible de ne pas relier ces concepts-limites à des concepts d'une autre connaissance ou de la connaissance générale.

2.2. Un exemple de BCT

Dans le cadre du laboratoire mixte ARAMIHS (Action Recherche et Application Matra/Irit en Interface Homme/Système), nous avons conçu un modèle de BCT que nous avons mis à l'épreuve des faits, d'abord à MATRA MARCONI SPACE puis, après qu'il eut été amélioré au sein de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), au CNES et à AÉROSPATIALE.

Les principales caractéristiques de ce modèle sont les suivantes.

- Une égale importance est accordée au fonctionnement linguistique (T) du terme (informations sur la nature, le genre, le type de complémentation...) et à son fonctionnement comme indice de concept (ou de concepts en cas d'homonymie) (C) (essentiellement, définition et relations avec les autres concepts).

- Les conditions d'usage (L T/C) constituent un champ à part entière ; elles s'intercalent dans le lien entre le terme comme mot d'une langue et comme indice de concept (en effet, c'est bien parce qu'on est dans telle ou telle entreprise, que l'on considère tel ou tel niveau d'expertise, que l'on peut justifier des divergences dans la dénomination). Cette organisation permet de rendre compte des cas courants où un même concept est dénommé par plusieurs termes dans une même entreprise, ce qui évite la normalisation trop brutale.

- On peut établir des relations entre les concepts ; ces relations, qui contiennent au moins est-une-sortre-de et a-pour-partie avec leur réciproque sont choisies en fonction du domaine.

Les relations entre termes sont calculées. Ainsi, deux termes seront totalement équivalents si :

- ils sont reliés au même concept ;
- ils ont la même nature grammaticale : en effet, nous considérons par exemple que, du point de vue conceptuel, un verbe et un nom, ou un adjectif et un nom... peuvent désigner le même concept, par exemple, "gel d'un test" et "geler un test" renvoient au même concept ; mais, si on parle d'équivalence linguistique, alors, l'appartenance à la même catégorie grammaticale est requise ;
- les valeurs respectives des attributs sont identiques ; si ces valeurs ne sont pas identiques, l'équivalence n'est pas totale.

Dans le schéma ci-dessous, T1 et T2 (qui renvoient au même concept) sont identiques si la catégorie grammaticale de T1 est la même que celle de T2 et si LT1/T2 et LT2/C1 sont identiques.

On peut, de la même façon, calculer l'équivalence interlinguistique en rajoutant une restriction sur la langue.

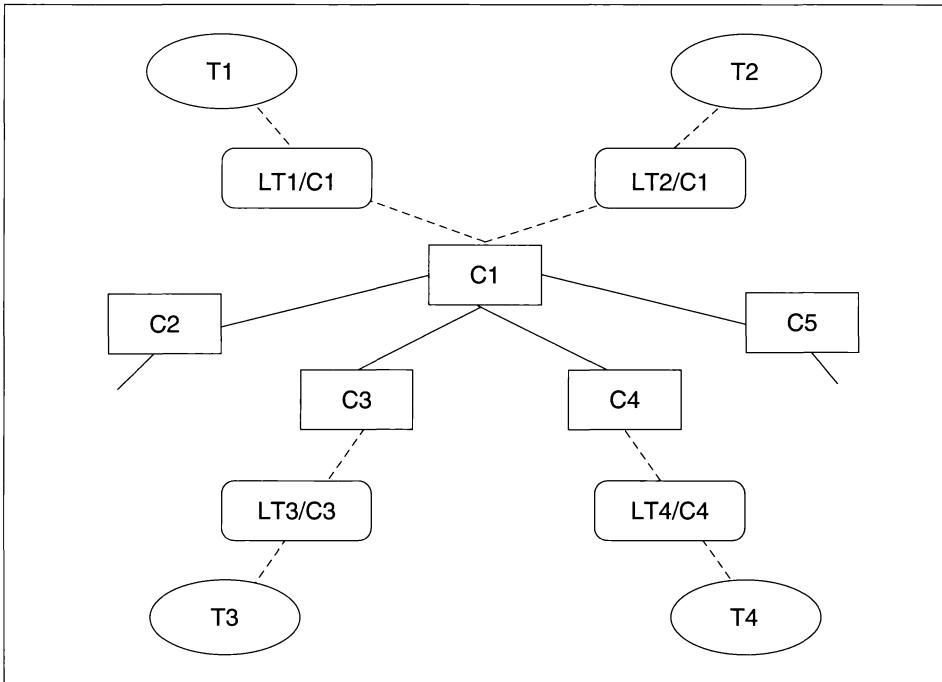


Figure 1 : Exemple de BCT

Les concepts C1, C2, C3, C4 et C5 sont reliés par des relations conceptuelles est-un, partie-de, éventuellement cause, conséquence... Chaque concept se manifeste par une réalisation linguistique (T) ou par plusieurs comme le concept C1. Les liens LT/C rendent compte des conditions d'usage.

La figure 2 propose un exemple de mise en œuvre du modèle.

Dans cet exemple, on a, pour le même concept étiqueté (arbitrairement) *capteur*, trois réalisations linguistiques : capteur, senseur et détecteur. Les conditions d'usage font intervenir l'utilisation effective dans l'entreprise et le statut dans une norme parue au *Journal Officiel* ; ces conditions d'usage n'étant pas identiques, les trois termes ne sont pas totalement équivalents bien qu'ils renvoient au même concept. Le concept *capteur* est intégré à un réseau de concepts dont une partie seulement est présentée ici : il est une partie du SCAO (Système de Contrôle d'Attitude et d'Orbite) et admet au moins deux concepts spécifiques : "capteur d'orientation" et "capteur de position".

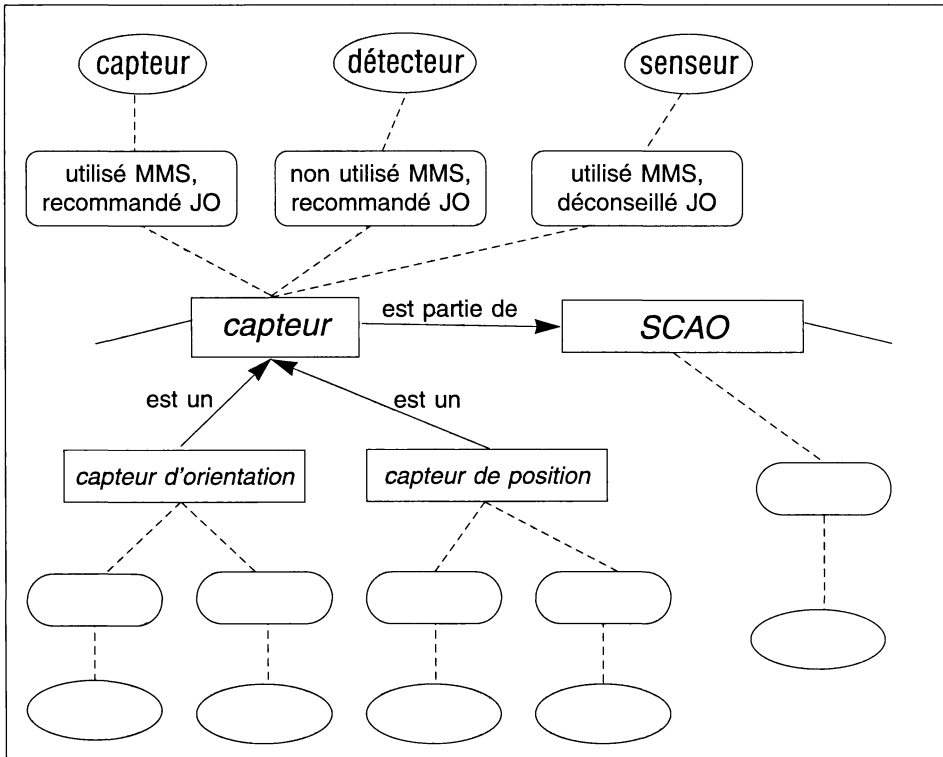


Figure 2 : Exemple de mise en œuvre de la BCT

2.2.1. Méthodologie de constitution de la BCT

La méthodologie de recueil et de traitement des termes que nous avons mise au point s'inspire à la fois des méthodes de sémantique des concordances et de celle de l'acquisition des connaissances de l'intelligence artificielle.

L'analyse linguistique se fait sur des gros volumes de données textuelles, à l'aide d'outils informatiques (extracteurs de termes candidats, outils d'analyse de contenu, concordanciers, gestionnaires de terminologie). Elle consiste à étudier le fonctionnement du terme et du ou des concepts auxquels il renvoie en analysant ses contextes d'apparition ; l'analyse se fait sur les deux plans, syntagmatique et paradigmatic (Condamines, 1992). De la même façon, une parenté de contextes aussi bien à droite qu'à gauche de deux termes différents amène à penser que ces deux termes renvoient en réalité au même concept. Enfin, l'étude des contextes permet d'établir une ébauche de réseau conceptuel.

L'entretien avec un (ou mieux, plusieurs experts) du domaine se fait sur la base des résultats de l'analyse linguistique. L'expert confirme ou invalide les

résultats et les complète, le cas échéant. La présence des textes permet une discussion objectivée, le linguiste n'étant là que pour rendre compte de ses conclusions ; l'expert peut ainsi, dans certains cas, être confronté à un fonctionnement (linguistique et conceptuel) qu'il ne connaissait pas, mais qui existe dans d'autres départements que le sien, ou auquel il ne pensait pas¹.

Dans la plupart des expériences de traitement des termes que nous avons, le linguiste est dans l'entreprise, sur les lieux même de production langagière, ce qui lui permet aussi de s'immerger dans la culture d'entreprise.

L'expérience de collaboration avec des entreprises, qui s'étale parfois sur plusieurs années, nous permet de rendre compte d'un très réel intérêt de leur part pour ce concept de Base de Connaissance Terminologique et pour sa mise en œuvre. Nous avons démontré qu'il est possible de construire de telles bases de connaissances. Il reste à prouver que l'originalité de cette approche est fructueuse, c'est-à-dire que ces BCT sont utilisées soit dans une consultation directe, soit par leur intégration dans des projets de plus grande envergure.

CONCLUSION

Notre expérience de collaboration avec des chercheurs en intelligence artificielle d'une part, avec des ingénieurs et des responsables de départements "Recherche et Développement" d'autre part, nous permet de rendre compte d'un réel besoin en entreprise et d'une réelle possibilité de rencontre entre les travaux sur la représentation des connaissances et les travaux sur la terminologie.

Cette rencontre n'en est qu'à ses débuts. Une réflexion approfondie reste nécessaire, elle devrait offrir un nouveau terrain d'investigation aux recherches sur les rapports entre langue et connaissances et sur les problèmes liés à la formalisation de la langue. Le besoin en entreprise devrait donner un coup de fouet à ces travaux en permettant aux chercheurs un accès facilité à la documentation technique et aux experts du domaine.

BIBLIOGRAPHIE

AHMAD K. (1993). Terminology and knowledge acquisition : A Text Based Approach. In K.-D. Schmitz (Ed.), *Proceedings TKE93 (Terminology and Knowledge Engineering)*. Frankfurt, Indeks Verlag, pp. 56-70.

AUSSENAC N., KRIVINE J.-P. & SALLENTIN J. (1992). Éditorial du numéro spécial sur l'acquisition de connaissance. *Revue d'Intelligence Artificielle*, vol. 6, n° 1-2, pp. 7-18.

1. La description de cette méthodologie est très succincte ; elle ne fait pas apparaître toutes les conditions préliminaires de sa mise en œuvre (choix du domaine, choix des relations conceptuelle, recueil des corpus...).

CONDAMINES A. (1992). Aide à l'acquisition de connaissance par la spécification de la terminologie d'un vocabulaire de spécialité. In *Actes des Journées d'Acquisition de Connaissance*. Dourdan, pp. 89-100.

CONDAMINES A. (1993). Un exemple d'utilisation de connaissances de sémantique lexicale : acquisition semi-automatique d'un vocabulaire de spécialité. *Cahiers de Lexicologie*, n° 62, pp. 25-65.

CONDAMINES A. & AMSILI P. (1993). Terminology between language and knowledge : an example of a terminological knowledge base. In K.-D. Schmitz (Ed.), *Proceedings TKE93 (Terminology and Knowledge Engineering)*. Frankfurt, Indeks Verlag, pp. 316-323.

CONDAMINES A. (à paraître). Terminology : New Needs, New perspectives. *Terminology*.

GAUDIN F. (1993). *Pour une Socioterminologie, des problèmes sémantiques aux pratiques institutionnelles*. Publications de l'Université de Rouen, n° 182.

KITTREDGE R. (1982). Variation and Homogeneity of Sublanguages. In R. Kittredge & J. Lehrberger (Eds), *Sublanguage. Studies of Language in Restricted Semantics Domains*. Berlin, New York, De Gruyter.

KOCOUREK R. (1982). *La langue française de la technique et de la science*. Wiesbaden, Brandestetter Verlag.

LERAT P. (1988). Terminologie et sémantique descriptive. *La Banque des mots*, numéro spécial, pp. 11-30.

LERAT P. (1989). Les fondements théoriques de la terminologie. *La Banque des mots*, numéro spécial, pp.51-62.

MEIJS W. & VOSSEN P. (1991). In so Many Words, Knowledge as a Lexical Phenomenon. In J. Pustejowsky & S. Bergler (Eds), *Lexical Semantics and Knowledge Representation, Proceedings Workshop 17 June 1991*. University of California, pp. 113-126.

MEYER I. (1991). Knowledge Management for Terminology-Intensive Applications : Needs and Tools. In J. Pustejowsky & S. Bergler (Eds), *Lexical Semantics and Knowledge Representation, Proceedings Workshop 17 June 1991*. University of California.

MEYER I., SKUCE D., BOWKER L. & ECK K. (1992). Towards a New Generation of Terminological Resources : an Experiment in Building a Terminological Knowledge Base. In *Proceedings of the Fifteenth International Conference on Computational Linguistics (COLING)*. Nantes, pp. 956-960.

PORTELANCE C. (1991). Fondements linguistiques de la terminologie. *META, Journal des traducteurs*, vol. 36, n°1, pp. 64-70. Presses Universitaires de Montréal.

RASTIER F. (1987). *Sémantique et Intelligence Artificielle*. *Langages*, n° 87. Paris, Larousse.

RASTIER F. (1993). *Sémantique et recherches cognitives*. Paris, PUF.

REY A. (1979). *La terminologie, noms et notions*. Paris, PUF, Que sais-je ?

SABAH G. (1988). *L'Intelligence Artificielle et le Langage*. Paris, Hermès.

SAGER J.C. (1990). *A practical Course in Terminology processing*. Amsterdam/Philadelphie, John Benjamins Publishing Company.

SKUCE D. & MEYER I. (1991). Terminology and Knowledge Engineering : Exploring a Symbiotic Relationship. In *6th International Workshop on Knowledge Acquisition for Knowledge-Based Systems*. Banff, Canada, pp. 21-29.

SLODZIAN M. (1995). La doctrine terminologique, nouvelle théorie du signe au carrefour de l'universalisme et du logicisme. In *Actes de Langue française et de Linguistique (ALFA)*. Universitas Dalhousiana, Halifax, Nova scotia, Canada, Vol. 7/8, pp. 121-136.