

Compte rendu

Ouvrage recensé :

Otman, Gabriel (dir) (1995) : « Terminologie et intelligence artificielle », *La Banque des mots*, Numéro spécial, Paris, Conseil international de la langue française, n° 7, 112 p.

par Patrick Drouin

Meta : journal des traducteurs / Meta: Translators' Journal, vol. 41, n° 4, 1996, p. 622-631.

Pour citer ce compte rendu, utiliser l'adresse suivante :

URI: <http://id.erudit.org/iderudit/017920ar>

DOI: 10.7202/017920ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

moins de 26 caractères du domaine *Typog.*, dont certains qui ne figurent pas le *Dictionnaire pratique* (caractère beurré, caractère brisé — ce dernier donné par le *Dictionnaire pratique* comme caractères allemands, dont l'équivalent serait *Fraktur* plutôt que *deutsche Schrift*).

Le dernier sondage, qui vise à se faire une idée de «la mise à jour très complète» signalée par les auteurs, porte sur le domaine de l'intelligence artificielle, mentionné nommément dans la préface. Ici, nous n'avons pas de dictionnaire français-allemand récent à notre disposition, et nous avons pris le *Vocabulaire des Technologies de l'information* (ISO / CEI JTC1 / SC1) publié par l'AFNOR en 1995 comme point de départ. Ce dictionnaire cite 26 sortes d'apprentissage (synonymes non compris), toutes faisant l'objet d'une norme ou d'une proposition de norme. Le *Compact* indique cinq sortes d'apprentissage marquées (IA), dont seul *apprentissage empirique* ne figure pas parmi les normes (on relève cependant *apprentissage heuristique*); en plus, *apprentissage automatique*, qui figure dans le dictionnaire spécialisé, ne comporte pas de marque de domaine dans le *Compact*. Les indications de numéro de normes ne sont pas fournies ici, car le domaine est sans doute trop récent, mais d'autres domaines comportent ce genre de citations.

Dernière remarque: le nombre de dictionnaires techniques français-allemand est relativement limité, et c'est la disponibilité des ouvrages qui a déterminé le choix de ces quelques brefs sondages. Il est permis de penser que le Kučera couvre des domaines entiers où il n'existe pas du tout de dictionnaire spécialisé pour la paire de langues français-allemand et ce n'est pas son moindre mérite.

JOHN HUMBLEY

CTN, Université Paris-Nord, Villetaneuse

■ OTMAN, Gabriel (dir.) (1995): «Terminologie et intelligence artificielle», *La Banque des mots*, Numéro spécial, Paris, Conseil international de la langue française, n° 7, 112 p.

L'engouement pour les approches à base de connaissances en terminologie ne cesse de se faire sentir depuis quelques années et il est donc tout à fait normal que les chercheurs qui s'intéressent de près ou de loin au mariage entre la terminologie et l'informatique soient tentés de sonder le terrain de l'*intelligence artificielle* (IA). Déjà, en 1992, un groupe de chercheurs sous la direction de Didier Bourigault (CAMS et EDF, Clamart) et d'Anne Condamines (ERSS-CNRS, Toulouse) considérait cette nouvelle avenue comme inévitable et mettait sur pied le *Groupe Terminologie et Intelligence Artificielle* (TIA).

À la veille de la tenue des 2^{es} *Rencontres Terminologie et Intelligence Artificielle* qui se tiendront à l'Université Toulouse-le Mirail sous la tutelle du Groupe TIA, ce numéro spécial de *La Banque des mots* (8 articles, 112 pages) vient nous remettre en mémoire les principales communications présentées lors de la première édition des rencontres T & IA. Cette première rencontre a su susciter l'intérêt de nombreux chercheurs (plus de 200 personnes) d'horizons différents comme la linguistique, l'informatique et,

comme le dit si bien Gabriel Otman (coordonnateur du numéro spécial), de la linguistique informatique et de l'informatique linguistique.

Comme toute science qui s'ouvre aux nouvelles avenues de l'IA, la terminologie, malgré sa naissance relativement récente, doit repenser ses bases théoriques. De par sa nature et ses affinités à jongler avec les représentations sémantiques, elle s'y prête bien. Le problème de représentation et de structuration des connaissances en vue d'applications informatiques demeure cependant entier et les chercheurs du domaine doivent tenter de déterminer l'information qui apparaîtra désormais dans les banques de terminologie traditionnelles qui se transforment peu à peu en bases de connaissances terminologiques afin de permettre le passage à l'ère de l'IA.

Dans sa contribution, Pierre Lerat (*Terme, mot, vocable*, pp. 5-9) s'intéresse au problème que pose l'objet d'analyse de la sémantique lexicale en terminologie. Doit-on étudier le *mot-signe*, le *vocable de discours* ou le *terme*? Selon lui, l'étude des termes (qui représentent la verbalisation des connaissances selon l'auteur) permettra de mettre en place des réseaux notionnels autour de notions clairement définies dans le cadre d'un texte thématiquement et socialement établi par la communication spécialisée. La construction de ces réseaux notionnels doit aussi reposer sur les relations classiques utilisées en sémantique lexicale : l'hyponymie, la synonymie, l'antonymie, etc.

Revisitant la naissance de la terminologie en tant que science, Monique Slodzian (*Comment revisiter la doctrine terminologique aujourd'hui?*, pp. 11-18) critique les assises actuelles de la méthodologie terminologique mise en place en grande partie dans le cadre de la *General Theory of Terms* (GTT) de l'école de Vienne. L'auteur explique la source de notions comme le *terme*, le *domaine* et, outil encore si cher aux terminologues il y a quelques années, l'*arbre de domaine*. L'empirisme logique à partir duquel ces notions ont été construites ne fait pratiquement plus partie de la pratique terminologique actuelle et nous devons donc nous demander s'il ne faut pas les réanalyser au moment où la terminologie tend à devenir une terminologie fondée sur les textes, ces banques de connaissances où le savoir scientifique se trouve exprimé à l'état brut. Le modèle aristotélicien de définition par genre prochain et différence spécifique est aussi critiqué par Slodzian, qui démontre son adéquation à l'aide de cas très simples. La sémantique cognitive permet de contourner ces problèmes fondamentaux à l'aide de modèles alternatifs de définition qui sont mieux adaptés aux réalités de l'IA par le recours aux prototypes et aux schémas. L'auteur rejoint Pierre Lerat sur l'importance de la prise en compte de la manifestation textuelle des termes qui sera un jour une source importante d'acquisition de connaissances aussi bien pour le traitement en terminologie qu'en IA. Les bases théoriques de la terminologie doivent donc être revues. On devra prendre en considération la dimension textuelle par le biais, entre autres, de la prise en compte du discours spécialisé et de ses diverses particularités (lexicales, sémantiques, syntaxiques, etc.).

Dans son article *Terme et dénotation*, Daniel Kayser (pp. 19-34) défend la thèse qu'en langue de spécialité, les termes ont pour fonction essentielle de désigner des sous-ensembles de l'univers de discours (construit et donné au départ par la spécialité) de façon permanente. À partir d'une étude du terme *champ* tiré d'un corpus de langue de spécialité (informatique), de l'analyse de ses différentes occurrences en discours et des glissements de sens qui s'en suivent, l'auteur parvient à démontrer que c'est dans le processus de compréhension de la phrase que se situe la dénotation d'un terme. Contrairement à la théorie qui veut qu'en langue de spécialité la capacité de dénomination soit épuisée par la définition, la fonction dénotative est, comme dans la langue usuelle, dérivée de la fonction inférentielle et non l'inverse.

Pour sa part, François Rastier (*Le terme : entre ontologie et linguistique*, pp. 35-65) décrit globalement la parenté entre ontologie et terminologie. Il aborde la problématique

des rapports du terme au non-terme, du terme au contexte et du terme au texte. C'est par une impressionnante démonstration, qui rejoint les propos des Daniel Kayser sur l'importance du texte dans la description terminologique, qu'il redéfinit le terme par le biais de la sémantique de textes scientifiques et techniques. Le recours, de plus en plus généralisé, que font les praticiens aux corpus, est un bon indice de ce que l'avenir nous réserve. La mise en place d'une véritable linguistique de corpus et une révision des positions théoriques de la doctrine terminologique ne peuvent que faciliter l'insertion de la terminologie dans le dynamisme créé par les travaux récents en IA.

À l'opposé de la thèse de Rastier, Bruno Bachimont (*Ontologie régionale et terminologie : quelques remarques méthodologiques et critiques*, pp. 67-86) soutient que la formalisation passe par la normalisation ou la décontextualisation des termes. Les systèmes à base de connaissances doivent savoir représenter la structuration des connaissances complètes d'un domaine plutôt que celle d'un expert. La définition du domaine devient ainsi critique et c'est, dans la majorité des cas, par le biais de la langue naturelle et de l'acquisition automatique des connaissances que les chercheurs doivent se tourner pour réussir à bien modéliser le domaine à décrire formellement. Il faut donc construire une théorie du domaine à partir de la langue naturelle et arriver ainsi à décrire les objets et leurs propriétés. Cette entreprise passe par la normalisation de la signification linguistique afin qu'il soit possible de leur attribuer un contenu prescriptif et conceptuel même hors contexte. Cette étape marque le passage d'une analyse linguistique à un système conceptuel cohérent, régulier et systématique permettant les inférences. Les notions qui composent le système se définissent comme une position ou une valeur dans le système des connaissances (et non de la langue). La normalisation sémantique, que l'auteur place dans le cadre du travail du terminologue, permet de construire un système conceptuel qui correspond à une ontologie du domaine. Cette dernière reçoit la dénomination *ontologie régionale* en tant qu'ontologie d'une partie du réel. Une étape supplémentaire est par la suite nécessaire pour le passage à une base de connaissances terminologiques, la détermination complexe d'une *ontologie computationnelle* qui ne saurait être identique à la première et qui ne saurait être autosuffisante.

Nadine Gros (*Extraction des énoncés avec le modèle pratique de KOD*, pp. 87-94) présente une génération d'outils KOD (*Knowledge Oriented Design*) destinés à l'acquisition et à la modélisation des connaissances. L'auteur démontre que ces outils peuvent intervenir dans le travail tant du linguiste que du cogniticien pour l'analyse du matériau que constituent les transcriptions d'entretiens avec les experts pour l'extraction des connaissances et leur structuration selon un modèle utilisable en IA. Le passage du niveau pratique (discursif) au niveau interne (modèle informatique des connaissances extraites) est présenté de façon à mettre en évidence l'interaction qui doit se créer entre le linguiste et le cogniticien afin de faire progresser les travaux de recherche qui amalgament terminologie et IA.

Le projet GALEN, présenté par Alan L. Rector (*Conceptual and linguistic knowledge in GALEN*, pp. 95-104), a pour objectif de mettre à la disposition de la communauté scientifique des outils pour la saisie, la représentation, la manipulation, l'échange et la diffusion de l'information clinique multilingue. Cela est rendu possible par divers outils qui permettent la description de la langue médicale et de sa terminologie par le biais d'un serveur terminologique qui utilise un modèle de description des concepts médicaux et de la terminologie fondé sur un langage formel propre à GALEN. Ce langage formel possède divers niveaux de connaissances dont le *niveau conceptuel* reposant sur une *interlingua* (visant une diffusion globale indépendante des langues et postulant une *interlingua* pouvant représenter l'information du domaine médical)

détachée à la fois de la langue naturelle (prise en charge par le *niveau linguistique*) et de l'inférence médicale. Les liens entre ces deux niveaux sont assurés par les *niveaux inférentiel, pragmatique et technique*. Un tel projet, sans une implication du terminologue spécialiste aux divers niveaux, ne saurait, à notre avis, être mené à bon port.

Les modèles de connaissances à objets présentés par Florence Lemaire et François Rechenmann (*Intégration de connaissances terminologiques dans les grandes bases d'objets: exemples en biologie moléculaire*, pp. 105-112), issus des langages informatiques orientés objet, se prêtent particulièrement bien à la description des objets ou des classes d'objets à l'aide de noms et d'attributs. Les modèles à objets autorisent une description hautement détaillée et structurée qui peut être intéressante pour une foule d'applications. L'utilisation d'un formalisme de description ne garantit cependant pas une description égale pour des travaux différents. Les bases de connaissances sont habituellement mises en place en vue de la résolution de problèmes bien précis et l'information ne peut que s'en ressentir. Il faut alors tenter de pallier ces problèmes de *subjectivité* des données en les reliant à des instances de texte explicatif ou justificatif. C'est à ce niveau que le lexique, ou la terminologie, peut avoir une grande influence en servant d'intermédiaire ou de pont entre la description formalisée des données et le matériau brut que constituent les textes. Les connaissances terminologiques semblent donc être un complément naturel aux connaissances formelles qui sont souvent trop rigides.

Cette première rencontre d'envergure réunissant les chercheurs en terminologie comme en IA a su ouvrir la voie à une réutilisation des connaissances qui ont été placées dans les banques de terminologie ainsi que de celles acquises par les spécialistes de la terminologie au fil des années. Espérons que la prochaine rencontre du Groupe TIA conduira à une fusion des énergies des chercheurs et à la mise en place de stratégies de recherche de plus en plus fructueuses.

PATRICK DROUIN
Université de Montréal, Montréal, Canada