

Pour une terminologie textuelle

Ce texte reprend de façon synthétique le tutoriel donné à l'ouverture des troisièmes journées *Terminologie et intelligence artificielle*.

Il présente l'analyse, faite par le groupe TIA, des nouveaux enjeux pratiques, théoriques et méthodologiques de la terminologie. Sur le plan pratique, l'accroissement des besoins en terminologie dans les entreprises et dans les institutions s'accompagne d'un élargissement qualitatif considérable de la gamme des produits à base terminologique nécessaires pour répondre à ces besoins. Ces évolutions entraînent des changements en profondeur de la pratique terminologique : l'activité de construction d'une terminologie est désormais essentiellement une tâche d'analyse de corpus textuels. Ils appellent du même coup à un renouvellement théorique de la terminologie : c'est dans le cadre d'une linguistique textuelle que doivent être posées les bases théoriques de la terminologie.

Termes-clés :
terminologie; théorie de la terminologie; terminologie textuelle; linguistique de corpus; ingénierie des connaissances.

1 Introduction

Ce texte constitue un résumé du tutoriel donné à l'ouverture des troisièmes journées *Terminologie et*

intelligence artificielle. Les points de vue présentés sont le résultat de discussions et réflexions menées depuis plusieurs années au sein du groupe TIA, qui, compte tenu de la complémentarité disciplinaire de ses membres et de leur expérience conjugulée dans le champ de la terminologie, est en mesure d'offrir une analyse suffisamment complète et équilibrée des nouveaux enjeux, pratiques, théoriques et méthodologiques, de la terminologie. Notons cependant qu'étant donné la nouveauté et la richesse qui sous-tendent cette nouvelle thématique, un certain nombre de questions sont toujours ouvertes et continuent d'être débattues au sein du groupe.

2 Applications de la terminologie : état des lieux

Les besoins en terminologie dans les entreprises et dans les institutions se multiplient. Suite à l'utilisation généralisée des outils de bureautique, à l'internationalisation des échanges, au développement d'Internet, la production de documents sous forme électronique s'accélère sans cesse. Or pour produire, diffuser, rechercher et exploiter ces documents, les outils de gestion de l'information ont besoin

de ressources terminologiques. Outre l'accroissement quantitatif de la demande, l'un des impacts essentiels de ces évolutions sur la pratique terminologique est l'élargissement qualitatif considérable de la gamme des produits à base terminologique nécessaires pour répondre à ces besoins. À côté des bases de données terminologiques multilingues classiques pour l'aide à la traduction, on voit apparaître de nouvelles productions terminologiques adaptées aux nouvelles applications de la terminologie en entreprise :

- Thésaurus pour les systèmes d'indexation automatique;
- Index structurés pour les documentations électroniques;
- Terminologies de référence pour les systèmes d'aide à la rédaction;
- Référentiels terminologiques pour les systèmes de gestion de données techniques;
- Ontologie pour les mémoires d'entreprise ou pour les systèmes d'aide à la décision;
- Réseaux lexicaux spécialisés pour les moteurs de recherche thématique sur le *Web*;
- Glossaires de référence et liste de termes pour les outils de communication interne et externe;
- Bases de connaissances terminologiques pour la description de corpus de référence;
- ...

L'essor de ces applications conduit à traiter, à l'aide d'outils, des quantités de documents considérables. Ce changement d'échelle met en évidence des phénomènes largement sous-estimés jusqu'ici.

C'est ainsi que s'impose le constat de la variabilité des terminologies : étant donné un domaine d'activité, il n'y a pas UNE terminologie, qui représenterait LE savoir sur le domaine, mais autant de terminologies que d'applications dans lesquelles ces terminologies ont été utilisées. Ces terminologies diffèrent quant aux unités retenues et à leur description selon l'application visée. Par ailleurs, la croissance terminologique, induite par la prolifération en tous sens de connaissances, entraîne la nécessité de mises à jour permanentes si l'on veut répondre aux besoins des utilisateurs.

Ce constat sur la variabilité remet en cause le principe de l'universalité des terminologies. L'expérience montre en effet qu'une terminologie élaborée pour une application à un moment donné n'est jamais identique à celle construite pour une application différente. Ces limites fortes à la réutilisabilité n'excluent pas des relations d'inclusion ou de chevauchements partiels entre terminologies dédiées à des applications différentes dans un même domaine d'activité.

L'ensemble de ces constats empiriques entraîne des changements en profondeur de la pratique terminologique : l'activité de construction d'une terminologie est désormais essentiellement une tâche d'analyse de corpus textuels. Ils appellent du même coup à un renouvellement théorique de la terminologie : c'est dans le cadre d'une linguistique textuelle que doivent être posées les bases théoriques de la terminologie.

3 Nouvelles pratiques terminologiques

L'activité de construction d'une terminologie devient avant tout une

tâche d'analyse de corpus textuels. Il y a à cela deux raisons essentielles :

- Les applications de la terminologie sont le plus souvent des applications textuelles (traduction, indexation, aide à la rédaction) ; la terminologie doit « venir » des textes pour mieux y « retourner ». C'est parce qu'elle n'est jamais déliée du texte qu'on parle de « terminologie textuelle ».
- C'est dans les textes produits ou utilisés par une communauté d'experts, que sont exprimées, et donc accessibles, une bonne partie des connaissances partagées de cette communauté, c'est donc par là qu'il faut commencer l'analyse.

L'expérience montre que l'hypothèse selon laquelle l'expert d'un domaine serait le dépositaire d'un système conceptuel qu'il suffirait de mettre au jour est non productive. La tâche d'analyse terminologique vise alors avant tout la construction d'une description des structures lexicales à l'œuvre dans un corpus textuel à partir d'une analyse réglée de ce corpus.

Cette tâche ne peut être menée à bien par les experts ; la médiation d'un analyste (linguiste terminologue, cogniticien) est nécessaire, en premier lieu parce qu'on colle trop à ses propres usages langagiers ; c'est le médiateur qui garantit la distance nécessaire à l'analyse. En second lieu, la pluralité des pratiques à l'intérieur de ce que l'on a coutume d'appeler « domaine » induit des points de vue différents sur le lexique (préférences, rejets, désaccords sur la définition) qu'il faut arbitrer. La division du travail linguistique à l'intérieur d'une entreprise requiert donc un médiateur, qui a en charge l'application.

Pour chaque unité choisie, l'analyste construit une signification (type) à partir des sens (occurrences) attestés dans le corpus. Dans cette tâche, il est guidé en amont par le corpus (spécificités lexicales) et en

aval par l'application (utilisation des descriptions).

L'expert doit être considéré comme un partenaire du linguiste terminologue, dans un travail de collaboration ; il est sollicité pour valider les descriptions construites par celui-ci.

Le domaine doit être lié à une pratique, maîtrisée par une communauté d'experts. Comme action (instrumentalisation du savoir propre à la technique), la pratique ne procède pas de connaissances statiques, liées à des expressions linguistiques bien stabilisées.

Avant la tâche de description lexicale, la constitution du corpus de référence est une étape essentielle, prise en charge par le linguiste terminologue. Il s'agit pour lui de collecter et de caractériser un ensemble de textes jugés pertinents pour l'application visée.

Devant la masse des données à analyser et les délais imposés, la tâche d'analyse de corpus ne peut être envisagée qu'avec l'utilisation des outils de la terminologie textuelle (concordanciers, extracteurs de candidats termes, extracteurs de relations candidates, classifieurs, etc.). L'utilisation de ces différents outils doit être encadrée par une méthodologie précisant à quel stade du processus et selon quelles modalités il convient de les utiliser.

4 Renouveau théorique

Ces changements en profondeur de la pratique terminologique appellent un renouvellement théorique.

Les propositions théoriques et méthodologiques qui suivent ont des bases empiriques ; elles sont issues d'une analyse des nouvelles pratiques de la terminologie, et elles ont pour ambition de les améliorer. Il ne s'agit

donc pas de fonder un nouveau dogme, mais de susciter des courants de recherche variés dans le champ de la linguistique, dont chacun pourra contribuer à cet objectif.

Proposition 1: objet empirique d'une linguistique textuelle, le texte est le point de départ de la description lexicale à construire. On va du texte vers le terme. Les bases théoriques de la terminologie doivent être ancrées dans une linguistique textuelle.

Proposition 2: le terme est un construit. Il est le produit d'un travail d'analyse, mené par le linguiste terminologue, dont les choix sont guidés par une double contrainte de pertinence :

- Pertinence vis-à-vis du corpus. Il s'agit de retenir et de décrire des structures lexicales qui présentent des caractéristiques à la fois spécifiques et stables. C'est à ce stade qu'intervient la validation par l'expert.
- Pertinence vis-à-vis de l'application. Les unités finalement retenues doivent l'être en fonction de leur utilité dans l'application visée, qui s'exprime en termes d'économie et d'efficacité. La validation est à chercher du côté des utilisateurs de l'application.

La tâche de description lexicale est un travail de fixation, de stabilisation, d'homogénéisation d'une signification, dont le résultat est le terme. Il s'agit de construire un type (une signification stable) à partir des occurrences manifestées en texte. C'est ainsi qu'on parle de *normalisation*, non plus au sens que la planification terminologique donne au mot, mais au sens où la communauté d'experts «entérine» des signifiés comme des termes du domaine.

Le résultat de la description peut se présenter sous des formes diverses : réseau, liste, glossaire, etc. Il n'existe pas de format canonique. Les noms ne sont pas les seules unités lexicales à décrire. En attribuant au terme la

fonction de dénommer les concepts, la terminologie classique privilégie les noms. En s'éloignant de cette approche référentielle très limitative, on est en mesure d'accueillir les autres catégories du discours (verbes, adjectifs, adverbes, prépositions, conjonctions), ainsi que des unités linguistiques plus ouvertes (syntagmes nominaux, verbaux, adjectivaux...).

5 Fin de la doctrine

Le virage méthodologique, rendu nécessaire par le travail sur corpus, crée une onde de choc qui ébranle les fondements de la doctrine wüsterienne, fortement référentielle (le mot comme étiquette du concept) et taxinomique (primauté de la relation générique/spécifique).

Il est illusoire de chercher à aménager la doctrine : le postulat d'une signification conçue comme discrète ou discrétisable, objectivante et permanente qui caractériserait le terme *a priori* est antinomique avec une terminologie textuelle. Les reformulations théoriques superficielles qui ont apparu ces dernières années sont vaines : la notion de «phraséologie», en particulier, ne peut sauver le postulat doctrinal du «terme» dans la mesure où elle est un biais pragmatique pour détourner la question du contexte et de l'unité terminologique.

Les termes ne sont pas des «unités de connaissances» qui viendraient «habiter la langue». La tâche d'analyse terminologique n'est donc pas un exercice de redécouverte d'un système notionnel préexistant qui caractériserait le domaine.

Les notions n'ont pas d'antériorité ou de priorité sur les mots : la terminologisation est un processus parallèle à l'élaboration conceptuelle.

La terminologie doit sortir d'une sémiotique du signe fondée sur la triade terme/concept/référent qui la rend inapte à aborder le texte. Cette critique du réductionnisme référentiel est à l'ordre du jour en philosophie du langage. Les appels à desserrer l'étreinte des postulats logicistes nous viennent de plusieurs côtés (Putnam, Auroux, Eco...), le positivisme logique qui a nourri la doctrine ayant été remis en cause dès la fin des années 60.

On peut constater par ailleurs que les connaissances nouvelles sont plutôt éphémères et partagées par des communautés restreintes au-delà desquelles elles ne circulent pas. On est loin de la conception idéalisée du domaine comme fragment de connaissances bien structurées, permanentes et clairement circonscrites.

On ne peut plus dire que la signification du terme est définie par la position du concept dans le système conceptuel correspondant dès lors que l'on met en doute la représentation métaphysique d'un système conceptuel préexistant représentable par l'arbre du domaine.

Il est aussi illusoire de se soumettre au référent, y compris dans les domaines techniques qui manipulent des artefacts. La description d'un objet technique est elle-même tributaire du point de vue imposé par la spécialité de l'expert. C'est en bout de chaîne, en normalisant le terme, qu'on lui prescrit une référence.

Dès que l'on abandonne l'approche logiciste du terme, étroitement liée à une sémantique véri-conditionnelle, on reconsidère le statut de la définition qui cesse d'être le résultat d'une procédure logique, métalinguistique. La définition doit être cohérente avec les sens contextuels (avérés en corpus) et pertinente vis-à-vis de l'application (comme elle s'inscrit dans une application, elle participe aux

objectifs communicationnels, elle doit être « localisée »).

Antinomique d'une approche étroitement onomasiologique, peau de chagrin du linguistique, l'approche textuelle ouvre largement les portes à tous les acquis de l'analyse linguistique et textuelle (on dépasse ainsi la vision étroite des LSP).

L'approche textuelle est descriptive (on analyse le fonctionnement d'unités lexicales en corpus) et non plus normative: les enjeux de la planification linguistique, si légitimes soient-ils, sont dissociés du travail terminologique proprement dit. L'objectif premier de la terminologie classique était la normalisation des langages techniques via la fixation *a priori* de la signification des mots. Les textes réels qui prolifèrent et circulent en tous sens, bousculant les frontières de domaines, remettent en cause ce projet de mise en ordre des termes *a priori*. Un tel programme de régulation prescriptive est contredit par le caractère fondamentalement ouvert des textes et de leurs signes. Le constat de la plasticité du donné linguistique conduit à refonder une « bonne pratique terminologique » sur le descriptif.

Pour conclure

L'actualité de la question terminologique au travers des changements intervenus en termes d'échelle et de rythme de production, ainsi que l'ampleur des besoins, appellent un renouveau théorique et méthodologique. En permettant d'aborder systématiquement l'étude des pratiques textuelles réelles, la linguistique de corpus, avec ses techniques et ses outils, donne accès aux expressions linguistiques concrètes d'où il sera possible de faire émerger, puis de normaliser les termes pertinents. C'est une formidable

ouverture pour la réflexion théorique et méthodologique. Il va sans dire que la question des procédures linguistiques présupposées par cette approche est à peine défrichée. Le groupe TIA entend participer au débat en pleine conscience de la complexité des enjeux théoriques et pratiques.

Il va également de soi que la linguistique ne peut couvrir à elle seule le processus complet de modélisation des connaissances; en fournissant la terminologie adéquate à l'application, le linguiste prépare le travail de représentation conceptuelle, mais il ne prend pas en charge la tâche de modélisation des connaissances qui aboutira à la construction d'une ontologie. Le relais est pris par l'ingénierie des connaissances. Le groupe TIA s'inscrit dans la nécessaire coopération interdisciplinaire entre linguistes et ingénieurs de la connaissance.

*Didier Bourigault,
Équipe de recherche en syntaxe
et sémantique,
CNRS,
Université de Toulouse Le Mirail,*

*Monique Slodzian,
Centre de recherche en ingénierie
multilingue,
Institut national des langues
et civilisations orientales,
Paris.*