

DIALANGUE

Volume 1

Avril 1990

BULLETIN DE LINGUISTIQUE

Module des lettres / Maîtrise en linguistique
Université du Québec à Chicoutimi

«[...] En ce qui concerne la langue, toutefois, nous possédons quelques certitudes. La première, dans l'ordre politique, c'est que la question de la langue est inflammable. La deuxième, c'est que qui perd sa langue perd son âme [...] La troisième, c'est que la langue doit être vue. Quand on ne voit plus sa langue, on cesse de la parler. Une ville saute d'abord aux yeux. Le problème de l'affichage est révélateur, c'est le cas de le dire. La quatrième, c'est que la langue doit être portée dans le coeur de ceux qui la parlent.»

Jean-Paul Desbiens,
«Sur le bout de la langue»,
La Presse (Montréal), 22 octobre 1986.

SOMMAIRE

- Les orientations de recherche des professeurs de linguistique de l'Université du Québec à Chicoutimi
 - Synthèse des mémoires de maîtrise
 - Mémoires et travaux de 1^{er} cycle
 - La situation du français à l'Université du Québec à Chicoutimi et dans la région
 - Comptes rendus d'ouvrages récents
-

ÉTUDE TERMINOLOGIQUE ET ANALYSE DES PROCÉDÉS DE FORMATION DE CINQUANTE NOTIONS EN TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES*

Martine Dubé

Dans une société axée sur les communications, l'implantation de nouvelles technologies pour accélérer le traitement de l'information s'avère inévitable. Ces technologies, mieux connues sous le nom de science informatique, regroupent plusieurs champs d'activités: que l'on pense par exemple aux logiciels, à la micro-informatique ou, tout simplement, à l'intelligence artificielle.

L'objet principal de l'intelligence artificielle réside dans la conception de programmes capables de reproduire certains aspects du raisonnement humain et tout particulièrement la compréhension d'une langue naturelle. Et, qui dit compréhension d'une langue naturelle pense tôt ou tard à cette faculté propre à l'être humain qu'est la communication.

C'est donc sur cette capacité de communiquer que se penchent les chercheurs en traitement automatique des langues naturelles qui travaillent à la mise au point de formalismes représentant la structure profonde d'une langue dans le but de permettre un jour la communication homme-machine. Pour ce faire, le langage utilisé par l'ordinateur doit se rapprocher la plus possible de celui que les humains parlent entre eux. Mais en plus d'avoir recours à la linguistique formelle et à la logique mathématique pour la conception de nouveaux formalismes reproduisant le langage humain, les spécialistes doivent se tourner vers les sciences cognitives et la psychologie pour tenter d'expliquer les mécanismes de compréhension qui se cachent derrière la communication.

La faculté de la parole sous-entend une certaine connaissance du monde qui permet au cerveau humain de faire les inférences nécessaires à la communication, dont l'ordinateur doit être doté pour communiquer efficacement.

Le traitement automatique des langues naturelles est une technique relativement nouvelle mais vouée à un essor considérable à cause des multiples applications qui en découlent: enseignement assisté par ordinateur, génération automatique de textes, traduction automatique, etc. C'est un domaine de connaissances qui se prête bien à une étude terminologique puisque les technologies de pointe, souvent aux prises avec un vocabulaire encore en formation, font face à des lacunes importantes relativement à la dénomination des nouveaux concepts.

* Résumé du mémoire de maîtrise en linguistique déposé à l'UQAC en février 1990.

L'objet de la terminologie est "[...] de repérer, d'analyser et, au besoin, de créer le vocabulaire pour une technique donnée [...] de façon à répondre aux besoins de l'utilisateur" (Dubuc cité dans Boulanger, 1985: 14). Avant que le vocabulaire ne soit définitivement fixé par l'usage, de nombreux cas de synonymie et d'emprunts sont rencontrés avec pour résultat un manque flagrant d'uniformité des termes utilisés pour dénommer un même concept. C'est au terminologue qu'il appartient de relever et de faire la lumière sur les termes à privilégier par l'intermédiaire des dossiers de terminologie qui constituent son principal moyen de transmettre à l'utilisateur les données recueillies. Car toute langue de spécialité a obligatoirement une terminologie propre qui est à la base de la communication.

En terminologie, la théorie et la pratique sont indissociables. La rédaction des dossiers qui constitue la partie pratique doit être soumise à un cadre méthodologique rigoureux. Tout aussi précise, la partie théorique consiste en une analyse du syntagme et du calque comme principaux modes de formation néologique retenus dans le cadre de notre recherche.

Le vocabulaire du traitement automatique des langues naturelles, à l'instar de celui de toute langue de spécialité, est composé de néologismes lexicaux ou termes nouveaux, en majorité des syntagmes, qui combinent plusieurs unités lexicales selon les règles morphosyntaxiques de la langue afin de dénommer les concepts nouvellement introduits en français. Le syntagme comme mode de formation, sous des allures bien françaises, a subi bien souvent l'influence de l'anglais puisque le concept est déjà dénommé dans cette langue; il ne suffit que de traduire les mots et à emprunter la forme en respectant la structure syntaxique du français. Mieux connue sous le nom de calque, cette pratique est en grande partie responsable, entre autres, de la présence massive de syntagmes dans les langues de spécialité.

En règle générale, on peut affirmer que le syntagme en traitement automatique des langues naturelles obéit à la structure fondamentale du français déterminé - déterminant proposée par Guilbert (1970 : 117) à l'exception des modèles empruntés intégralement à l'anglais.

Parmi les cinquante-neuf syntagmes du corpus, on en dénombre six types. Le modèle nom + adjectif est de loin le plus fréquent avec ses trente et une occurrences. "L'adjectif apparaît, en effet, comme l'instrument le plus apte à fournir des réalisations particulières et différenciées d'unités lexicales préexistantes" (Guilbert 1965 : 270). Cette combinaison produit des syntagmes simples (**grammaire discontinue**) et des syntagmes complexes (**grammaire discontinue contextuelle**). Ainsi, que le premier élément soit une lexie simple (**grammaire**) ou une lexie complexe (**grammaire discontinue**) il occupe la fonction de déterminé ou base. Le second élément, l'adjectif dans le présent modèle, joue le rôle de déterminant.

Certaines bases comme **analyse** et **grammaire** sont dotées d'une vocation de génération à cause de l'étendue de la notion qu'elles couvrent et donnent lieu à des séries d'unités terminologiques répondant à ce modèle. L'adjectif, en tant que second élément du syntagme, permet non seulement de constituer des unités de signification nouvelles dans un champ sémantique nouveau, mais permet de distinguer, et parfois même d'opposer, les unités lexicales issues d'une même base par leur trait le plus significatif (**grammaire usuelle**, **grammaire**

sémantique). Dans le cas des syntagmes complexes, l'adjectif permet de restreindre le contenu sémantique de la base comme c'est le cas pour **grammaire discontinue** qui devient **grammaire discontinue contextuelle** et **grammaire discontinue statique** par l'adjonction d'une seconde détermination qui vient ajouter un élément distinctif aux syntagmes ainsi formés.

Le modèle **nom + joncteur + nom** est le second en importance avec seize occurrences. Au lieu d'être établi par juxtaposition comme avec l'adjectif, le rapport syntagmatique est obtenu par liaison par le biais d'une préposition. Les expansions de la base, qu'elles soient introduites par la préposition **de**, **à** ou **en**, jouent le même rôle que l'adjectif, mais leur degré de cohésion varie en fonction de la préposition utilisée. La préposition **de** assure le plus fort degré de cohésion et de lexicalisation comparable à celui de l'adjectif parce qu'en liant la base à son déterminant nominal, elle confère au nom la même fonction que l'adjectif. C'est d'ailleurs celle qu'on rencontre le plus souvent avec onze occurrences (**grammaire de métamorphose**, **réseau de transition ...**), suivie du **en** avec trois occurrences (**analyse en série**, **analyse en parallèle**, **interface en langue naturelle**) et du **à** avec deux occurrences (**extraposition à gauche** et **grammaire à clauses définies**).

Responsable de la présence massive de syntagmes dans les langues de spécialité, le calque est le second mode de formation néologique en importance en traitement automatique des langues naturelles. Sur un total de quatre-vingt-douze termes, on compte soixante calques complets ou partiels et une dizaine d'emprunts intégraux. De ce nombre, il y a quatre propositions de remplacement: **grammaire discontinue contextuelle**, **grammaire des cas**, **grammaire discontinue statique** et **grammaire lacursive**. On appelle proposition de remplacement tout terme proposé par le terminologue qui vise à remplacer un terme existant qui lui paraît plus ou moins satisfaisant du point de vue de la langue. Les cinquante-six calques restants, puisés à même la documentation dépouillée, sont issus des milieux spécialisés.

Parmi les six types de calques contenus dans le corpus, la **traduction littérale** ou mot à mot d'un syntagme (**précédence linéaire** pour *linear precedence*) ou d'un sigle (**GE** pour *XG*) d'une langue à l'autre est la plus courante avec vingt-sept occurrences. Dans le premier cas, le terme **précédence** a été calqué de toute pièce sur l'anglais et constitue un néologisme formel en français. Le terme **préséance** est l'équivalent français proposé qui paraît plus acceptable.

Vient ensuite l'**adaptation syntaxique**, avec quinze occurrences, qui respecte dans la séquence **déterminé - déterminant** les rapports logiques qui unissent ces éléments au moyen de prépositions (**grammaire de case**, **principe de compositionnalité ...**).

À la lumière des résultats obtenus, il est permis de constater que l'anglais en traitement automatique des langues naturelles, tout comme en informatique et en intelligence artificielle, influence grandement les modes de formation du vocabulaire. Mais la maîtrise du calque et de ses mécanismes apparaît comme un mode de formation rapide et efficace, le recours massif au syntagme en est un exemple signifiant, et un moyen de limiter la pénétration de l'anglais dans les langues de spécialité qui sont en constante évolution.

Bibliographie

BOULANGER, Jean-Claude (1985), Initiation à la terminologie. Notes de cours et travaux pratiques, Québec, Université Laval, août.

GUILBERT, Louis (1965a), "Les cadres morphologiques du néologisme dans la formation du vocabulaire de l'aviation", dans La formation du vocabulaire de l'aviation, chapitre 5, Paris, Librairie Larousse, p. 255-329.

GUILBERT, Louis (1970b), "Théorie du néologisme", dans Cahiers de l'Association internationale des études françaises, no 5, mai, p. 9-29.