

Pour une nouvelle terminographie

Gilles Bélanger

Volume 36, numéro 1, mars 1991

La terminologie dans le monde : orientations et recherches

URI : <https://id.erudit.org/iderudit/002047ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

[Découvrir la revue](#)

Citer cet article

Bélanger, G. (1991). Pour une nouvelle terminographie. *Meta*, 36(1), 49-54.

POUR UNE NOUVELLE TERMINOGRAPHIE

GILLES BÉLANGER
Université de Montréal, Montréal, Canada

INTRODUCTION

Depuis une vingtaine d'années, une petite armée de terminologues a mis dans les mains de divers utilisateurs, en particulier des traducteurs, un nombre impressionnant d'ouvrages lexico-terminologiques de toutes sortes. Presque sans exceptions, ces ouvrages visent à recenser tous les termes ayant un rapport quelconque avec un «domaine» plus ou moins nettement délimité, tentant en cela de respecter un précepte théorique selon lequel la terminologie doit viser à l'exhaustivité. C'est ainsi que dans différents domaines de pointe, des ouvrages ont acquis une immense réputation, certes bien méritée, en multipliant les mises à jour et les entrées. Les dictionnaires bilingues de H. Piraux et de M. Ginguay en sont des exemples par excellence. Mais en même temps, on peut dire, pour reprendre en la biaisant un peu une terminologie de terminologues, que les auteurs de terminologies ont jusqu'ici préféré travailler en extension plutôt qu'en compréhension: ils cherchent à recenser le plus grand nombre possible d'unités terminologiques obtenues par des combinaisons de mots différents, plutôt qu'à analyser les caractères des éléments constitutifs de ces unités.

Cette tendance a sa contrepartie dans l'utilisation que l'on fait des ouvrages terminologiques bilingues et multilingues. Il n'est pas un traducteur, par exemple, qui ne soit atteint d'une forme malgré tout assez bénigne du syndrome de *Dunkin' Donuts*. De cela on se rend bien compte dans l'enseignement de la traduction scientifique et technique; devant une difficulté de traduction, le premier réflexe du traducteur n'est plus de chercher à décomposer le terme complexe et à en analyser les constituants compte tenu du contexte, mais à se croiser mentalement les doigts et à se dire: «*Il faut qu'il y soit!*»

La langue de l'informatique m'est apparue comme un choix tout à fait approprié pour illustrer ce phénomène. Les observations qui suivent pourraient toutefois s'appliquer à bien d'autres langues de spécialité, à quelques nuances près.

Après une période initiale au cours de laquelle l'emprunt et la création néologique spontanée ont joué un rôle central dans la création de la langue informatique, cette dernière crée aujourd'hui de nouveaux termes en faisant appel principalement à deux modes privilégiés: la composition par expansions déterminatives et la siglaison.

Ce dernier procédé constitue en fait une tentative de ramener des termes complexes à des termes simples, d'autant plus facilement mémorisables qu'ils peuvent être assimilés à de nouveaux termes mono- ou bisyllabiques. Dans ce domaine, la langue informatique est passée maître. Dans un monde envahi par les sigles, il est évident que le facteur mnémonique joue un grand rôle dans la création. C'est ainsi que des entreprises d'avant-garde se sont lancées dans la production de micro-ordinateurs basés sur des processeurs de type RISC (Reduced Instruction Set Computer), opération très risquée dans cette jungle de la compatibilité. EISA n'est pas une ensorceleuse, mais une nouvelle norme de bus (Extended Industry Standard Architecture), élaborée par la bande des neuf pour

concurrer *Big Blue*. D'autres sigles sont devenus de véritables monogrammes qui donnent aux disciplines qui les possèdent leurs lettres de noblesse, comme en témoigne le prestige rattaché à CAD (Computer-Aided Design) et à CAM (Computer-Aided Manufacturing). Aujourd'hui, les lettres de l'alphabet n'étant qu'au nombre de vingt-six, c'est la bataille rangée dans certains quartiers. Ainsi, au moins trois disciplines se disputent le créneau CAT: Computer-Aided Testing, Computer-Assisted Training et Computer-Assisted Translation. Que la meilleure gagne! Les modes sont toutefois passagères. Après l'ésotérisme, après l'élitisme, il faudra trouver un nouveau «siglo-moteur»!

On pourrait discourir longtemps sur les sigles et leurs modes de formation, mais mon propos vise plutôt à attaquer le mythe de l'exhaustivité. On peut en effet se demander si l'exhaustivité est un objectif atteignable, compte tenu du mode même de formation des termes techniques, qui tend à privilégier des matrices terminogéniques N + Adj, N + Prép + N (Portelance 1989: 400) et N + N pour le français, et Adj + N et N + N (et diverses variantes intégrant, par exemple, un participe passé adjectivé, tel *computer-assisted teaching*) pour l'anglais. Les lois de permutation montrent qu'un nombre relativement petit d'unités (moins d'un millier) peut engendrer un nombre extrêmement grand (plusieurs millions) de combinaisons, certes pas toutes valables linguistiquement, mais indicatives d'un ordre de grandeur du possible. On peut donc supposer que les entrées d'un ouvrage terminologique quelconque, quelle que soit sa richesse, ne forment qu'une petite partie du vocabulaire potentiel de la discipline traitée. Afin d'illustrer mon propos, j'ai confronté le *Dictionnaire anglais-français d'informatique* de M. Ginguay (1987) à un corpus de textes divers traitant d'informatique.

ANALYSE D'UN CORPUS

L'échantillon était constitué de neuf textes en langue anglaise (cf. annexe) de 1 000 mots chacun, choisis dans une gamme allant de la vulgarisation à la recherche fondamentale. Ces 9 000 mots ont produit 368 termes simples et complexes reliés au domaine; le tableau 1 donne leur répartition, ainsi que celle d'un échantillon systématique portant sur 11 % du dictionnaire de Ginguay. La longueur moyenne des termes était de 2,08 mots pour le corpus, contre 1,86 mot pour le Ginguay. Ces chiffres semblent indiquer une sous-représentation, par rapport aux textes du corpus, des termes complexes de plus de deux mots dans le Ginguay.

TABLEAU 1 — Répartition des termes dans le corpus et dans Ginguay

| | 1 mot | 2 mots | 3 mots | 4 mots + | longueur moyenne |
|---------|-------|--------|--------|----------|------------------|
| Corpus | 29,9 | 42,9 | 19,8 | 7,4 | 2,08 |
| Ginguay | 33,7 | 49,9 | 13,4 | 3,0 | 1,86 |

Le taux de réponse du Ginguay, c'est-à-dire le pourcentage des termes du corpus trouvés tels quels dans le dictionnaire, a varié considérablement selon le type de texte, allant de 73 % pour un texte de vulgarisation à 10 % pour un article de recherche (figure 1) Le Ginguay a fourni en moyenne une réponse sur deux, avec un taux de réponse considérablement plus élevé pour les termes de un et de deux mots (52 %) que pour les termes de trois mots et plus (15 %).

Je n'ai pas cherché à comparer le Ginguay avec d'autres dictionnaires d'informatique, celui-ci ayant été reconnu *a priori* comme un bon dictionnaire bilingue pour ce domaine de spécialité. Comment alors expliquer ce taux relativement faible de réponse? Un début d'explication peut se trouver, comme le souligne Geneviève Mareschal (1989 : 378), dans la «concurrency ambiguë existant entre les formes complètes et les formes réduites». L'exemple qu'elle donne concerne justement l'informatique : à côté de la forme complète *automatic text formatting feature*, elle relève dans le même texte quatre formes réduites désignant la même réalité : *automatic text formatting*, *text formatting feature*, *automatic formatting*, *text formatting*. Ces cinq termes renferment en tout quatre mots différents seulement, dont trois de la langue générale.

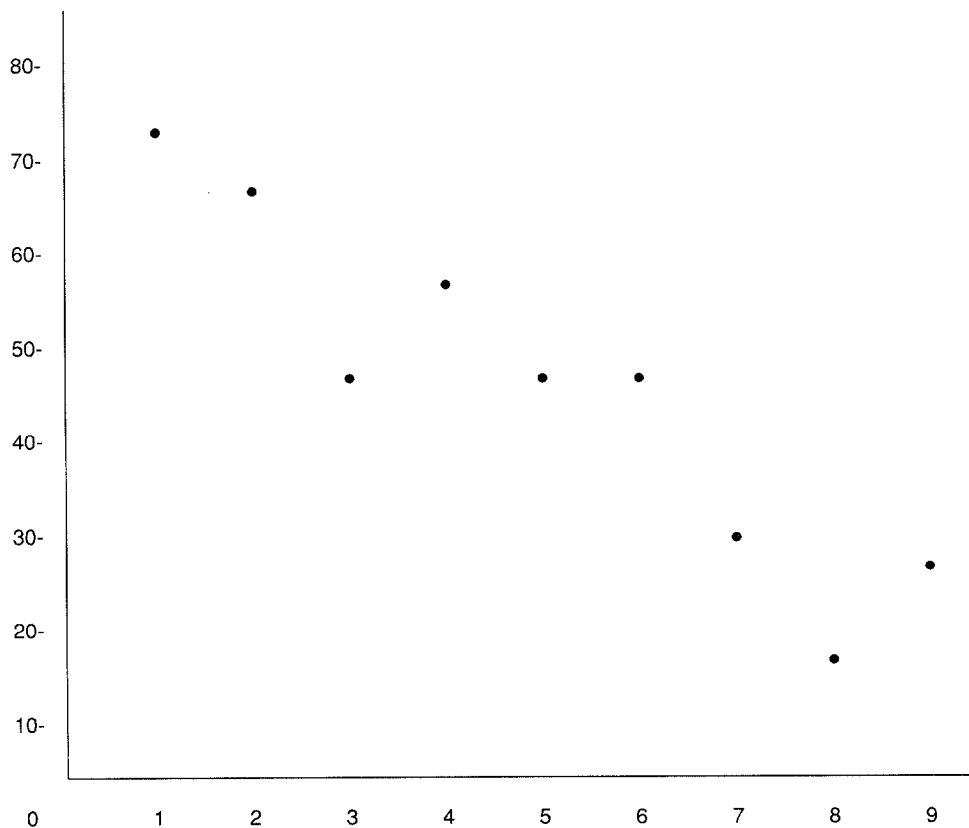


FIGURE 1 – Taux de réponse du Ginguay dans un échantillon de 9 textes.

Type des textes

1. Article de vulgarisation
2. Article de revue pour décideurs
3. Article de revue technique pour public averti
4. Article de revue technique professionnelle
5. Publicité destinée aux professionnels
6. Ouvrage didactique technique
7. Ouvrage didactique universitaire
8. Monographie
9. Article de revue pour chercheurs

En tentant toutefois de relever le taux de réussite à partir des constituants des termes complexes, je me suis heurté à un certain nombre de problèmes. Par exemple, y a-t-il une nuance de sens entre *graphic* et *graphics*, dans des termes comme *graphic animation system* et *noneditable graphics page preview*, ou encore entre *communication* et *communications*, *application* et *applications*, etc. L'article «control», qui occupe plus d'une page dans le Ginguay et dont la structure est inchangée depuis les premières éditions, n'est pas fait pour fixer les idées sur les différents sens de ce terme. Dans la deuxième partie de l'article, censée représenter le sens de «vérification» du terme anglais, on compte plus d'une vingtaine d'équivalents rendus par *commande*. On pourrait croire que tout cela est expliqué dans la version unilingue française du dictionnaire de Ginguay. Nenni. Toutes ces nuances y sont expédiées en une dizaine de lignes. Mais si vous avez par bonheur la chance de traduire une expression du type *computer-aided something*, 21 exemples différents vous indiqueront, comme si ce n'était pas encore assez clair, que cela se traduit par «...assisté par ordinateur». Si le terme à traduire renferme le mot «architecture», vous ferez sans grand risque un pari sur son sosie français, en vous disant que la langue informatique n'en est pas à un emprunt près, mais que vous auriez bien aimé pouvoir en saisir le sens exact.

Ces quelques remarques n'ont pour but que de grossir certains défauts des terminologies bilingues et d'en faire ressortir certaines lacunes, principalement en ce qui a trait au contenu notionnel. L'objectif visé, mais jamais atteint, est l'exhaustivité, et cela aux dépens de l'analyse notionnelle.

AUTRES DONNÉES STATISTIQUES

Quelques données de la statistique lexicale permettront de faire certaines autres constatations. Les trois premières colonnes du tableau 1 sont tirées de Christine Portelance (1987: 357); elles concernent cinq glossaires analysés dans le cadre d'un projet de terminotique.

TABLEAU 2 — Données de l'analyse de cinq glossaires selon Portelance (1987)

| Nbre de termes | Nbre de mots | Nbre de mots différents | Mots/terme | Mots diff./terme |
|----------------|--------------|-------------------------|------------|------------------|
| 374 | 776 | 424 | 2,07 | 1,13 |
| 1 128 | 2 875 | 1 010 | 2,55 | 0,9 |
| 2 129 | 5 246 | 1 667 | 2,46 | 0,78 |
| 3 291 | 9 243 | 2 232 | 2,81 | 0,68 |

Ces données montrent que le nombre de mots différents par terme tend à diminuer à mesure qu'augmente le nombre de termes du glossaire. Cela est dû à la redondance, qui se manifeste aussi bien dans les terminologies spécialisées que dans les textes suivis, et qui conduit à une saturation du vocabulaire, phénomène mis en évidence par Jean Baudot (1986) (Portelance 1987: 357). (Curieusement, le nombre de mots par terme tend, lui, à augmenter avec le nombre de termes, mais il n'est pas possible de formuler quelque hypothèse que ce soit à partir des données brutes.) Par ailleurs, toujours selon Portelance, pour un même glossaire, le pourcentage de mots différents (c'est-à-dire de graphies différentes) diminue fortement à mesure qu'augmente le nombre de termes, passant d'un peu moins de 50 % pour une tranche de 500 termes à un peu plus de 20 % pour environ

10 000 termes. La saturation du vocabulaire que révèlent ces chiffres serait encore plus accentuée (10 % dans l'hypothèse de l'auteure) s'ils avaient été établis pour des vocables lemmatisés. Si l'on prend comme exemple le dictionnaire de Ginguay (1987), qui allègue 13 600 entrées, cela voudrait dire environ 1 360 vocables différents.

On peut considérer que ces quelques centaines de mots constituent la base notionnelle du domaine de l'informatique. Ces mots se répartissent dans au moins trois grandes catégories :

- les mots tirés de la langue courante et qui n'ont pas besoin de longues explications pour en faire comprendre le sens général. Ainsi de *automatic*.
- les mots du vocabulaire général des sciences et des techniques, qui ont une charge sémantique relativement stable, mais qui peuvent présenter certaines particularités suivant le domaine d'emploi. Pensons à des mots comme *rate*, *density* ou *level*, mais aussi à *control*, à *input*, etc.
- les mots qui constituent le noyau propre au domaine, qui sont en général monosémiques dans les limites du domaine.

Les combinaisons de mots de ces trois catégories devraient pouvoir s'analyser à partir des éléments de sens de chacun des mots de la combinaison, compte tenu toutefois d'un certain nombre de restrictions. Ainsi, un nombre relativement restreint de combinaisons, dues surtout à des créations spontanées tirant parti de procédés terminogéniques moins analytiques qu'analogiques (par exemple, *électron célibataire* en chimie-physique, ou *widow/orphan protection* en traitement de texte), ont une charge sémantique qui n'est pas la somme de la charge de chacun des constituants. De même, un nombre tout aussi restreint de termes complexes, constitués par des expansions déterminatives du type *hybrid local network architecture* peuvent être sources d'ambiguïtés. Toutefois, une connaissance minimale du domaine, ou une brève recherche, permet de repérer dans ces syntagmes complexes des constituants de termes complexes, par exemple *local network*, *hybrid local network* et *network architecture*.

CONCLUSION

Parallèlement à une terminographie s'exerçant en extension, il serait intéressant de développer une autre pratique, axée celle-là sur la définition des concepts. À défaut de fournir des équivalences toutes faites, cette pratique aurait l'avantage immense de mettre en évidence

- les relations lexico-sémantiques entre constituants de termes;
- les éléments de sens stables et récurrents des termes complexes d'un même domaine;
- les différents sens et emplois des mots polysémiques.

Cette pratique aurait aussi l'avantage d'encourager étudiants et traducteurs à pratiquer une terminologie «dynamique», s'adaptant au contexte, et de laisser l'utilisateur des ouvrages terminologiques la liberté de décoder le message, une fois acquises les notions essentielles.

BIBLIOGRAPHIE

- GINGUAY, M. (1987) : *Dictionnaire anglais-français d'informatique*, 9^e éd., Paris, Masson.
- MARESCHAL, Geneviève (1989) : «Repérage d'unités terminologiques dans le contexte de l'enseignement de la traduction spécialisée», *Meta*, 34-3, pp. 377-380.
- PORTELANCE, Christine (1987) : «Fertilisation terminologique ou insémination terminologique artificielle?», *Meta*, 32-3, pp. 356-360.
- PORTELANCE, Christine (1989) : «Syntagmes et paradigmes», *Meta*, 34-3, pp. 398-404.

Annexe

1. Niklaus WIRTH, «Data Structure and Algorithms», *Scientific American*, 251-3, Sept. 1984, pp. 60-62.
2. Gerry BLACKWELL, «OA's New Frontier», *OA (Office Automation)*, 5-1, Jan. 1989, pp. 16-18.
3. Edward MENDELSON, «Two Aces and a King», *PC Magazine*, 8-20, Nov. 28, 1989, pp. 97-98.
4. Dennis LINNELL, «The 9370 Attraction», *PC Tech Journal*, 5-11, Nov. 1987, pp. 175-177.
5. «Introducing the AST Premium/386», publicité.
6. Rodney ZAKS and Austin LESEA, *Microprocessor Interfacing Techniques*, 3rd ed., Berkeley, Calif., Sybex, 1979, pp. 377-383.
7. David KROENKO, *Database Processing*, 2nd ed. Chicago, SRA, 1983., pp. 205-208.
8. Edward YOURDON and Larry L. CONSTANTINE, *Structured Design: Fundamentals of a Discipline of Computer program and Systems Design*, Englewood Cliffs, N.J., Prentice-Hall, 1979, pp. 331-333.
9. Nick BURD, «CSMA/CD Microcontroller Communication Network for Low-level Control», *Microprocessors and Microsystems*, 13-7, Sept. 1989, pp. 427-428.