

Syntagmes et paradigmes

Christine Portelance

Volume 34, Number 3, septembre 1989

1. Actes du Colloque Les terminologies spécialisées : Approches quantitative et logico-sémantique et 2. Actes du Colloque Terminologie et Industries de la langue

URI: <https://id.erudit.org/iderudit/003858ar>

DOI: <https://doi.org/10.7202/003858ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Portelance, C. (1989). Syntagmes et paradigmes. *Meta*, 34(3), 398–404.
<https://doi.org/10.7202/003858ar>

SYNTAGMES ET PARADIGMES

CHRISTINE PORTELANCE
Université McMaster, Hamilton, Canada

Le corpus qui sert à notre étude est constitué de 10 300 termes français tirés du Vocabulaire Électrotechnique International de la CEI¹, vocabulaire multilingue de termes normalisés. Cette liste de termes a été informatisée à l'Université de Montréal dans le cadre du projet BATEM² mené par le GRESLET³.

Lors de sa conception, le principal objectif de BATEM était l'optimisation du stockage et de la structure interne des données. Cette optimisation a été obtenue à partir de l'étude de la saturation du vocabulaire. Personnellement, je m'intéresse plutôt à la saturation du vocabulaire en tant que manifestation d'un phénomène morphosyntaxique propre aux nomenclatures. Toutefois, avant d'aborder cette question, j'aimerais rappeler certaines conclusions du projet BATEM, lesquelles ont été diffusées par Baudot (1987)⁴.

LA SATURATION DU VOCABULAIRE

Baudot fait observer qu'en considérant les termes comme du matériel textuel et les mots dont ils sont formés comme le vocabulaire du texte, on peut établir un rapport de similarité entre la saturation du vocabulaire d'une liste de termes spécialisés et celle observée dans des textes continus. En effet, Baudot constate que le nombre de mots graphiques croît avec le nombre de termes mais que cet accroissement est de plus en plus faible au fur et à mesure que le nombre de termes augmente : le pourcentage de mots différents est de 33,8% pour 1 000 termes français et il tombe à 15,2% pour 10 300 termes : ce taux de saturation est comparable à celui de textes continus en français. Baudot attribue cet état de faits à trois facteurs : 1. certaines catégories de mots sont privilégiées, 2. les mots qui composent les termes sont puisés à même un réservoir limité. 3. les syntagmes terminologiques ont une formation syntaxique plus figée.

En effet, les termes de cette liste sont constitués à 83% de syntagmes : les constituants de ces termes se combinent selon certains modèles syntaxiques et des constituants identiques se retrouvent dans de nombreux termes. Cette présente étude examinera surtout les données relatives à la répétition de constituants et à leurs combinaisons, plutôt que les données concernant la description des configurations syntagmatiques qui sont au nombre de 386 et qui méritent à elles seules toute une étude.

Si on retire du vocabulaire les mots vides (présentés à la figure 1), le taux de saturation des 10 300 termes passe de 15,2% à 20,2%.

Une fois les mots vides retirés, les cent mots les plus fréquents (0,02% des mots différents) correspondent à 6,820 occurrences, soit 30% des occurrences. On y retrouve huit adjectifs : *électrique*, *magnétique*, *automatique*, *électronique*, *acoustique*, *continu*, *inverse* et *radioélectrique*. Il y a également non qui peut être soit adverbe, soit nom comme dans *circuit NON*. Sur les 91 substantifs restants, 57 sont à la fois termes et constituants de termes, ce sont, par ordre de fréquence :

Figure 1

Mots vides		
<i>de:</i> 3003	<i>du:</i> 118	<i>avec:</i> 42
<i>d':</i> 1470	<i>l':</i> 115	<i>aux:</i> 26
<i>à:</i> 1175	<i>des:</i> 102	<i>dans:</i> 23
<i>un:</i> 317	<i>ou:</i> 70	<i>le:</i> 21
<i>en:</i> 304	<i>au:</i> 57	<i>sans:</i> 20
<i>par:</i> 288	<i>pour:</i> 53	<i>entre:</i> 18
<i>une:</i> 132	<i>et:</i> 44	<i>les:</i> 18
<i>la:</i> 126	<i>sur:</i> 43	<i>avant:</i> 11
Total des occurrences: 7596		Total de formes diff.: 24

courant (304); *tension* (255); *puissance* (162); *contact* (136); *signal* (121); *durée* (100); *charge* (96); *commande* (95); *électrode* (95) *relais* (95); *élément* (93); *rayonnement* (90); *résistance* (89); *enroulement* (88) *réseau* (79); *antenne* (76); *action* (69); *réponse* (68); *champ* (67); *énergie* (65); *protection* (64); *compteur* (62); *dose* (61); *convertisseur* (60); *bruit* (59); *enregistrement* (59); *erreur* (59); *potentiel* (57); *câble* (55); *émetteur* (55); *émission* (55); *arc* (53); *service* (52); *tarif* (50); *sortie* (49); *caractère* (47); *phase* (47); *position* (47); *sensibilité* (46); *programme* (45); *cathode* (44); *entrée* (44); *réglage* (44); *bobine* (43); *connecteur* (42); *régulation* (42); *soudage* (42); *rendement* (41); *code* (40); *terre* (39); *couplage* (38); *polarisation* (38); *commutateur* (37); *connexion* (37); *diffusion* (37); *écran* (37); *capacité* (36); *support* (36).

La liste suivante présente les mots qui ne sont que constituants :

temps (1868); *système* (171); *dispositif* (108); *appareil* (107); *caractéristique* (102); *effet* (102); *coefficient* (78); *essai* (70); *bande* (68); *constante* (59); *courbe* (56); *fréquence* (53); *ionisation* (53); *chambre* (50); *détecteur* (50); *rapport* (47); *angle* (46); *distorsion* (46); *boîte* (45); *densité* (44); *poste* (44); *source* (44); *composante* (42); *lecture* (42); *retour* (41); *ligne* (39); *section* (39); *surface* (38); *espace* (37); *état* (37); *taux* (37); *pouvoir* (36).

Pour atteindre 50% des occurrences, il faut ajouter à ces mots les 223 mots suivants dans la liste de fréquence, ce qui veut dire que 50% des occurrences sont produites par 0,07% des mots du vocabulaire. Les mots qui n'apparaissent qu'une fois constituent 49,3% des mots différents. Le tableau des fréquences, comprenant les mots vides, est présenté à la figure 2.

La similarité des phénomènes de saturation de vocabulaire entre les textes continus et les nomenclatures tient à des raisons différentes. Si la concordance des chiffres existe, la nature des répétitions ainsi que la nature des mots répétés est différente. Sur le plan de

Figure 2

Fréquences	Nombre de formes différentes
+500	3
499-200	5
199-150	4
149-100	13
99-80	9
79-60	16
59-60	18
49-40	32
39-30	46
29-25	38
24-20	72
19-15	104
14-10	158
9	67
8	69
7	87
6	143
5	153
4	224
3	342
2	745
1	2273

la communication, cette similarité est liée aux mécanismes de redondance inhérents au fonctionnement de tout code ; du point de vue terminologique, la redondance est un procédé économique.

REDONDANCE ET COMBINATOIRE

Dans l'étude de cette liste de termes, on doit considérer la redondance lexicale comme une combinatoire. Une distinction doit être faite entre la réutilisation de mots et la réutilisation de termes ; dans ce deuxième cas, il n'y a pas seulement répétition de monoterme mais également répétition de multitermes. Pour illustrer ces types de répétitions, je ne dispose pas de statistiques sur l'ensemble des termes mais seulement sur 6 867 termes, soit 67% du corpus, ce qui constitue cependant un échantillon suffisant pour tirer des conclusions valables.

Sur 6 867 termes, 1 040 termes, soit 15%, sont à la fois termes et constituants de termes, parmi ceux-ci, près de la moitié (504) sont des syntagmes. Les deux types de constituants les plus fréquents sont, comme on pouvait s'y attendre, [N + AJ] (187 cas) et [N + PR + N] (184 cas) totalisant 73,6% des syntagmes termes et constituants. Des monoterme servent donc à former des multitermes qui à leur tour peuvent également en

former d'autres. La redondance lexicale est due en majeure partie à la présence de séries dans la nomenclature. On distingue deux types de séries : les séries formées par les syntagmes à noyaux identiques et celles formées par les déterminations communes à plusieurs termes. Certains termes font partie des deux catégories de séries. Comme exemple, examinons la répartition des occurrences du mot *courant*, le constituant de termes le plus fréquent.

LES SÉRIES AVEC COURANT

Parmi les 304 occurrences du mot *courant*, 167 sont des têtes de syntagmes. Cette série de syntagmes à noyaux identiques est formée de 628 mots dont 178 formes différentes pour un pourcentage de saturation de 28%. Ce pourcentage est élevé compte tenu du nombre de termes, il est comparable au taux de saturation observé dans une tranche de 2 000 termes de cette même liste de termes.

Le modèle [N + PR + (DT) + N] engendre 64 termes, soit plus du tiers de la série et il connaît des expansions, comme dans les exemples suivants :

1. *courant de court-circuit*
2. *courant de court circuit permanent*
3. a *courant de court-circuit monophasé permanent*
- b *courant de court-circuit biphasé permanent*
- c *courant de cout circuit triphasé permanent*

Par l'ajout de déterminations au terme *courant*, on obtient les termes 1, 2 et 3 qui désignent des notions en relation subordonnée tandis que les notions désignées par les termes 3.a, 3.b, et 3.c sont en relation coordonnée.

Le modèle [N + AJ] engendre 35 termes et connaît également des expansions :

1. *courant présumé*
2. *courant présumé établi*
3. *courant présumé coupé*
4. *courant présumé symétrique*

Dans l'analyse des termes 2, 3 et 4, sans l'existence du terme 1, on peut être tenté d'appliquer la même analyse que dans : *un individu présumé innocent* c'est-à-dire : [*courant [présumé établi]*], mais si l'on tient compte d'un système de relations paradigmatiques propres à la terminologie, il faut plutôt privilégier l'analyse : [[*courant présumé*] *établi*], ce qui peut apparaître surprenant au premier abord. Reprenons l'analyse avec le terme 4 : *courant présumé symétrique*. Le courant n'est pas présumé... symétrique, il est +[présumé] et +[symétrique]. Et si on devait donner un ordre de formation aux termes, l'existence de la série nous amène à poser la formation de *courant présumé* comme précédant celle de *courant présumé symétrique*. On peut appeler ce type d'analyse : analyse en constituants immédiats déterminés par pertinence terminologique, ce qui correspond 2 à peu de choses près⁵ à ce que Baudot et Clas (1982)⁶ ont nommé : l'identification de sous-expressions significatives de terme (SXS).

On peut observer qu'il y a également combinaison de déterminations différentes, soit les déterminations : [adj] et [prép + N], comme dans les exemples suivants :

1. *courant minimal de commande*
2. *courant minimal de maintien*
3. *courant critique de grille*
4. *courant inverse d'électrode*
5. *courant élémentaire de conduction*

Dans les trois exemples, il y a insertion d'un adjectif dans des termes déjà existants : *courant de commande*, *courant de maintien*, *courant de grille*. À l'exemple 4, une détermination [PR + N] est ajoutée au terme *courant inverse*. Le dernier cas est une fusion de deux termes : *courant élémentaire* et *courant de conduction*. Tous ces termes doivent être considérés comme des SXS (sous-expressions significatives) des termes donnés en exemple.

On ne peut représenter sous forme hiérarchique tous les rapports qu'entretiennent entre eux les syntagmes ayant comme noyau courant, mais il n'en reste pas moins qu'ils sont tous reliés à la même notion mère dont le nom sert de genre. La série de syntagmes avec comme noyau courant s'explique si on examine cette série d'un point de vue notionnel, c'est-à-dire du point de vue de la théorie scientifique à partir de laquelle se sont développées les industries électriques et électrotechniques.

«Si une théorie physique n'admet rien de semblable au courant électrique, un petit nombre de critères, qui peuvent varier considérablement d'un cas à l'autre, suffiront à identifier les courants, même s'il n'y a aucun ensemble de règles spécifiant les conditions nécessaires et suffisantes de son identification.»⁷

Si une théorie physique n'est pas à même de fournir les règles d'identification des notions maîtresses, la formulation de règles sémantiques, comme l'utilisation d'un système de traits distinctifs, s'avèrera d'autant plus difficile. On est ici confronté aux limites de la fonction classificatrice que l'on prête souvent aux nomenclatures techniques. Or, l'approche classificatrice, qualifiée de possible dans le cas des nomenclatures techniques, est jugée difficile voire impossible par Alain Rey (1979)⁸ dans le cas des nomenclatures scientifiques. La série *courant* offre ici un bel exemple de la quasi impossibilité de départager clairement ce qui appartient en propre au domaine technique et ce qui appartient au domaine scientifique, d'où l'usage de plus en plus répandu de l'expression *technoscientifique*.

Par ailleurs, les déterminations sérielles avec *courant* peuvent prendre différentes formes : [prép + N] comme dans *entrée de courant*, *prise de courant*; [N + adj] comme dans *convertisseur de courant alternatif*, et *convertisseur de courant continu*. Le terme *courant continu* est utilisé comme détermination dans trente syntagmes et *courant alternatif* est présent dans 19 syntagmes. Ils peuvent être liés au reste du syntagme par différentes positions : *à*, *de*, *en*, *par*, et *pour*. Il y a également d'autres formations qui n'entraînent pas nécessairement des séries : [N + N] comme dans *caractéristique tempscourant* et [N + Adj + Adj] comme dans *caractéristique de courant coupé limité* ou *courant coupé* est un terme ainsi que *courant coupé limité*. Tous les multitermes formés avec *courant* et utilisés comme déterminations n'entraînent pas nécessairement des déterminations sérielles.

SYNTAGMES ET PARADIGMES

Un texte dont le noyau est *courant* ne reçoit sa signification que de l'existence de d'autres termes de la série et comme nous l'avons vu, les différents *courant* sont distingués les uns des autres sans que la théorie physique ne fournisse de réelle définition du *courant électrique* lui-même. Si on avait à nommer un nouveau *courant*, il faudrait tenir compte de la série, puisque les séries forment des paradigmes. En effet, les termes entretiennent entre eux des rapports paradigmatiques qui ne sont pas sans influencer leur formation. De plus, il existe deux types de rapports paradigmatiques dans une nomenclature : le rapport paradigmatique direct, celui qu'un terme entretient avec les termes de la série de syntagmes à noyau identique à laquelle il appartient, et le rapport paradigmatique oblique, celui qu'un terme entretient avec les termes d'une autre série lorsqu'une sous-expression significative de ce terme appartient à une autre série. Par exemple, le terme *déclencheur à maximum de courant* aura un rapport paradigmatique direct avec les termes de la série *déclencheur* et un rapport paradigmatique oblique avec les termes de la série *courant*.

De l'examen de ces données, il me faut conclure qu'on ne peut étudier la syntagmatique des termes, les rapports entre les constituants, sans tenir compte de l'existence de séries, c'est-à-dire sans tenir compte de l'axe paradigmatique. La longueur d'un syntagme terminologique par exemple, peut être conditionnée par le nombre et le type de rapports paradigmatiques entretenus avec les autres termes. À première vue, on peut penser qu'une configuration de 14 éléments (dont cinq substantifs) n'apparaissent qu'une seule fois dans le corpus (0,00019%) des termes) devrait être considéré comme une sorte de résidu sans réelle importance dans l'étude de la formation des termes. Or, sur les 293 configurations de substantifs 52,2% n'apparaissent qu'une seule fois ; il semble alors difficile de les mettre de côté sans tenter de les expliquer. Si ces longs termes semblent être des phénomènes d'exception lorsqu'on les analyse en fonction de l'axe syntagmatique, leur caractère bizarroïde s'estompe la plupart du temps lorsqu'ils sont examinés par rapport à l'axe paradigmatique. Ces termes existent parce qu'ils sont formés de termes et parce qu'ils s'inscrivent dans des séries.

L'existence d'une combinabilité des termes peut également servir à expliquer les différences de figement que l'on observe généralement entre les composés de la langue générale et ceux des langues de spécialité. Un terme désigne bien une seule notion, mais puisque certains de ses constituants peuvent être des termes qui, combinés à d'autres termes, sont susceptibles d'en engendrer d'autres, le lien entre de tels constituants sera senti comme plus lâche. Cette combinabilité constitue le potentiel de dénomination d'une nomenclature.

L'INFORMATISATION DE DICTIONNAIRES

En envisageant ces statistiques d'un point de vue pratique, on peut affirmer que si dans un domaine technoscientifique on devait nommer une nouvelle notion, il y a de fortes probabilités pour que le nouveau terme soit un syntagme et qu'au moins un des constituants soit présent dans le vocabulaire de la nomenclature du domaine. C'est pourquoi, il faut absolument considérer le potentiel de dénomination des nomenclatures lors de l'informatisation de dictionnaires technoscientifiques. Il faut rendre ce potentiel transparent : on devrait pouvoir afficher les séries existantes en distinguant les deux types de rapports paradigmatiques de manière à fournir à l'utilisateur — à défaut du terme cherché — tout au moins les stratégies linguistiques de dénominations. Un tel dictionnaire pourrait alors être considéré comme un véritable produit de la terminotique.

NOTES

1. Commission Électrotechnique Internationale
2. BATEM : banque de terminologie fonctionnant sur micro-ordinateur
3. Groupe de recherche en sémantique, en lexicologie et en terminologie de l'Université de Montréal
4. Lors des Premiers Entretiens du Centre Jacques-Cartier, au Colloque «Les moyens informatiques au service de la langue française», Lyon, 1987.
5. Les auteurs proposent une analyse en constituants, mais sans le critère paradigmatique de la pertinence terminologique que je propose maintenant. Dans cet article, les cas d'ambiguïté syntaxique sont tranchés par la connaissance du domaine.
6. «De la naissance des 'banquettes' ou un modèle de minibanque de terminologie bilingue», *La Banque des mots*, Paris, CILF, 1982.
7. Kuhn, John (1979) : *La structure des révolutions scientifiques*, Champs, Paris, Flammarion.
8. Rey, Alain (1979) : *La terminologie noms et notions*, Que-sais-je ? no. 1780, Paris, PUF.