

L'arbre de domaine en terminologie

Massiva N. Zafio

Volume 30, numéro 2, juin 1985

URI : <https://id.erudit.org/iderudit/004635ar>

DOI : <https://doi.org/10.7202/004635ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Zafio, M. N. (1985). L'arbre de domaine en terminologie. *Meta*, 30(2), 161-168.
<https://doi.org/10.7202/004635ar>

L'ARBRE DE DOMAINE EN TERMINOLOGIE

La terminologie , en dépit de sa relative jeunesse, est une discipline mûre, tendant, déjà, vers le statut de science. De ce fait, elle comporte une méthodologie mettant en jeu un certain nombre de méthodes et d'outils de travail, notamment l'**ARBRE DE DOMAINE**.

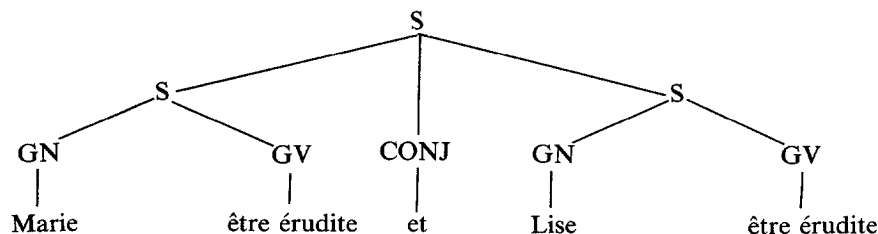
Grosso modo, cet arbre, encore appelé **ARBRE ANALOGIQUE** , est la représentation, sous une forme pyramidale, des notions clés d'un domaine et des relations qu'elles entretiennent entre elles. Ce type de représentation n'est pas, cependant, propre à la terminologie. D'autres disciplines en font usage. Ces arbres ont pour noms : organigramme, structure arborescente, arbre généalogique et thésaurus.

L'organigramme est l'arbre utilisé en administration pour représenter les différents paliers hiérarchiques d'une organisation donnée. Il est trop bien connu pour s'y attarder.

« Structure arborescente » est le nom, vague certes puisqu'il peut se rapporter à toute représentation d'idées sous la forme d'un arbre, que les linguistes ont donné à cet

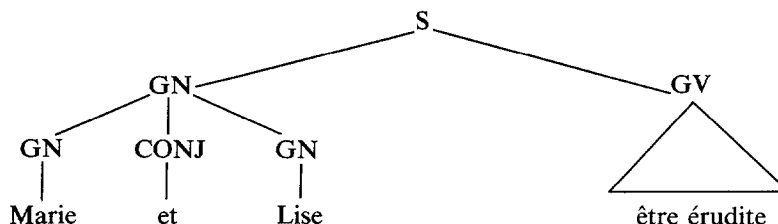
arbre qui leur permet de représenter certaines structures phrastiques. Ainsi, la phrase : « Marie et Lise sont érudites » se présente selon le schéma suivant :

Figure 1.



Cette structure peut se synthétiser comme suit :

Figure 2.



GN	= groupe nominal	S	= phrase
GV	= groupe verbal	CONJ	= conjonction

Les deux arbres décrits plus haut sont, néanmoins, assez éloignés de l'arbre des terminologies. Ils présentent, avec ce dernier, une lointaine ressemblance de forme : ils sont pyramidaux, et de structure : le sommet de la pyramide représente une entité, et la base, les parties.

La généalogie, on le sait, est la science qui a pour objet la recherche de l'origine et de la filiation des familles. L'instrument de cette recherche est l'arbre généalogique que Norbert Hézelles et Nadine Vigier (1979 : 134) définissent comme étant un « tableau donnant, sous la forme d'un arbre avec des nominations, la filiation des membres d'une famille ». À l'origine, cet arbre était le dessin d'un arbre véritable avec tronc, branches, feuilles et fruits. Puis, pour des raisons de commodité, on en est venu à une figure dont la réalisation était plus aisée.

Il en existe de deux sortes, selon que la démarche généalogique privilégiée est ascendante — l'ancêtre est en bas du schéma, près des racines, si l'on veut — ou descendante — l'ancêtre est au sommet de l'arbre.

Avec l'arbre généalogique, nous nous approchons de l'arbre analogique : les généalogies sont hiérarchisées ; les membres d'une même génération, telles des notions de même extension, sont situés sur le même plan horizontal ; les générations sont reliées entre elles par un trait de filiation, comme le terme spécifique, dans un arbre de domaine, est lié au générique par un trait de liaison. Nous reviendrons plus en détail sur ces aspects.

Avec le thésaurus, nous entrons de plain-pied dans l'arbre de domaine. C'est, selon nous, le père de l'arbre que nous utilisons en terminologie, vu l'antériorité de la bi-

bliothéconomie par rapport à la terminologie. Du thésaurus, mot latin signifiant « trésor », Marie-Thérèse Laurielle (1981 : 3-4) donne cette définition :

Du point de vue de sa forme, le thésaurus est un vocabulaire organisé, comportant des termes ayant entre eux des relations sémantiques et génériques, qui s'applique de façon exhaustive à un domaine de la connaissance.

Plus loin, elle précise :

Du point de vue de sa fonction, (c'est) un instrument de terminologie utilisé pour traduire en un langage artificiel moins libre (un arbre), le langage naturel employé dans les documents par les indexeurs ou les utilisateurs, ainsi que pour revenir au langage naturel à partir du langage artificiel.

De ces deux citations nous pouvons isoler deux points :

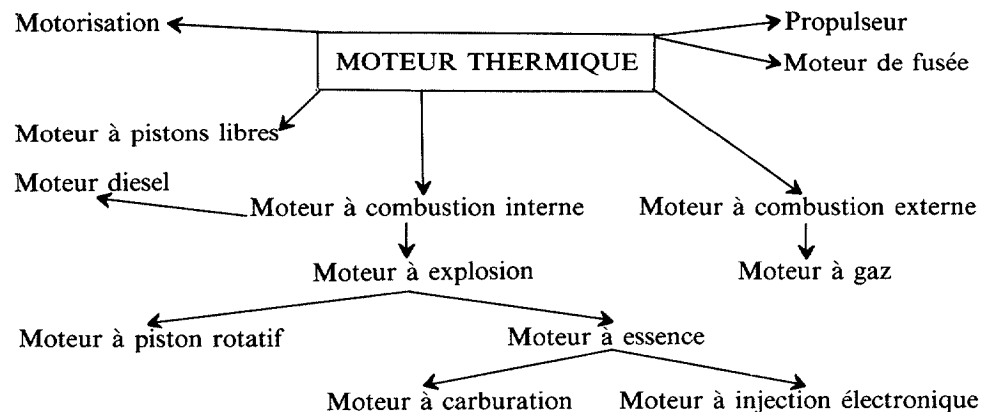
◆ *Le thésaurus est un langage artificiel.* C'est en effet une façon figurée, donc artificielle, de représenter des liens notionnels entre des mots du langage naturel que nous utilisons.

◆ *Le thésaurus est un instrument de terminologie* puisqu'il représente un effort de structuration, d'organisation du vocabulaire d'un domaine de connaissance. C'est donc une nomenclature arborescente.

Il peut revêtir plusieurs formes dont les plus courantes sont : le schéma fléché et le schéma circulaire.

Dans le schéma fléché, le sujet qui justifie la figure constitue le point central d'où rayonnent différentes sous-catégories. Ainsi, un schéma fléché représentant les moteurs thermiques peut être élaboré de la façon suivante :

Figure 3.



Le sujet de la figure 3, le « moteur thermique » se détaille en plusieurs sous-catégories, qui, à leur tour, peuvent engendrer des catégories plus fines.

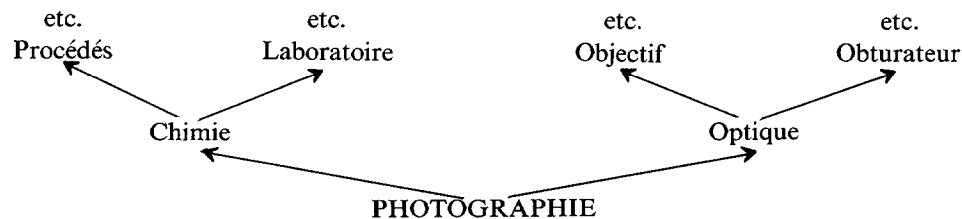
L'autre type de thésaurus, tout aussi courant mais mieux organisé, est le schéma circulaire. Il se présente sous la forme de cercles concentriques ou d'un cône, selon le point de vue de l'observateur. Le centre du plus petit cercle (le sommet du cône) porte la case contenant le sujet de la figure. De là, rayonne un premier niveau des sous-catégories qui occupent l'orbite la plus rapprochée du centre. De ces sous-catégories émanent d'autres catégories qui se situent sur une orbite légèrement plus éloignée, etc.

La preuve n'est donc plus à faire quant à la filiation entre le thésaurus et l'arbre de domaine. En effet, ici et là, nous avons un objet en commun : le vocabulaire d'un domaine. L'un et l'autre utilisent des relations allant des catégories génériques vers des catégories plus spécifiques car plus fines ; enfin, ici comme là, on parle de liaisons sémantiques et génériques. Une description de l'arbre de domaine et, surtout, de sa structure interne nous convaincra davantage.

Il existe, en terminologie, deux types d'arbres de domaine : l'*arborescence verticale* ou à *progression verticale* et l'*arborescence horizontale* ou à *progression horizontale*. Nous les décrirons tour à tour.

L'arborescence verticale est un arbre, comme l'illustrent les figures ci-dessous et dont le sommet se situe au point le plus haut du schéma. Il rappelle l'arbre généalogique en généalogie descendante. À l'instar de celui-ci, nous l'avons vu plus haut, l'arbre de domaine admet une variante : c'est l'arbre que propose Robert Dubuc dans son *Manuel pratique de terminologie*. Comme l'arbre en généalogie ascendante, cet arbre vertical a son sommet au bas de la figure et progresse vers le haut.

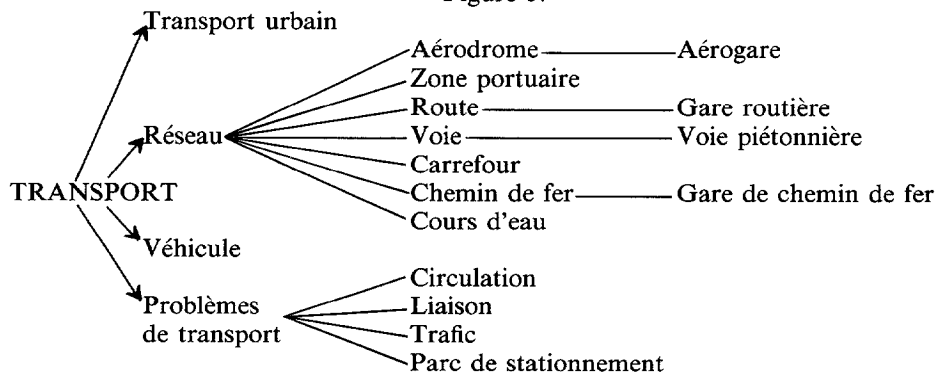
Figure 4.



Très rare d'emploi, il est difficile à réaliser car la progression s'effectue comme à reculons. L'élaboration commence au bas de la page et se termine vers le haut, démarche inverse du processus normal de lecture dans lequel les yeux balaient les lignes supérieures du texte et descendent, ligne après ligne, vers le bas.

Dans l'arbre de domaine à progression horizontale, le sommet de la figure occupe l'espace le plus à gauche de la page. Nous en avons une illustration chez Auger et Rousseau in *Méthodologie de la recherche terminologique*, p. 18. Il se présente ainsi :

Figure 5.



Source : Daumais et Payen (1975) : *Étude pour un traitement automatique des sources en histoire des techniques*, Paris, CRAM.

Cet arbre, quoique très commode puisqu'il se dispose peut-être mieux que le précédent, comporte un inconvénient : sa progression est limitée par la longueur de la page, limitation qui peut contraindre le terminologue à omettre certains embranchements. Il convient, cependant, aux domaines plus étendus.

Il est à noter que ces arbres ne sont pas l'exclusivité de la seule terminologie : les documentalistes et les généalogistes les utilisent aussi.

Tous ces arbres que nous avons examinés sont intéressants du point de vue de leur forme. Ils le sont davantage quant à leur structure et leur mode de fonctionnement.

L'analyse de la structure interne des arbres de domaine révèle deux catégories de relations : relations entre les cases ou les descripteurs au plan horizontal et les relations au plan vertical.

En lecture horizontale, les cases présentent la caractéristique d'entretenir entre elles une relation d'identité de catégories ou de parallélisme. Expliquons. Tous les descripteurs appartenant à un même niveau constituent des notions de la même catégorie, c'est-à-dire que tous les mots du même niveau doivent avoir la même extension. Ainsi, au niveau 2 de la figure 6 en annexe, nous avons des types de moteurs : moteurs à piston, turbine à gaz. Au niveau 3, nous avons des groupes de propulsion.

Autre particularité des cases en lecture horizontale : elles sont mutuellement exclusives. Ainsi, sur la même figure (en 2.1 et 2.2), « moteurs à piston » exclut « turbines à gaz ».

Niveau 4 : « avec éléments tournants et « sans éléments tournants » s'opposent » ;

Niveau 8 : « réacteur double-flux avec mélange à l'éjection et « réacteur double-flux sans mélange à l'éjection » s'exclut ».

En plan vertical, on peut identifier deux types de relations : les relations génériques et les relations partitives.

Une relation générique est une relation d'inclusion de termes au sens réduit, c'est-à-dire spécifiques, dans des termes couvrant une aire sémantique plus grande. Vues verticalement, les notions de niveaux inférieurs sont incluses dans celles des niveaux supérieurs. Un terme générique représente une classe de concepts dont fait partie le concept représenté par le terme spécifique. Ce dernier diffère du concept générique par au moins une caractéristique. On peut de ce fait, considérer les notions de niveaux supérieurs comme des catégories, et les notions de niveaux inférieurs comme des sous-catégories, c'est-à-dire des spécialisations de traits sémantiques que recèlent les catégories supérieures. Ainsi, le mot « turbine à gaz » comporte plusieurs traits sémantiques, plusieurs éléments de signification ; c'est un moteur à réaction qui peut entraîner soit une hélice, soit un rotor d'hélicoptère, ou encore fonctionner de façon pure. Ce sont ces différents traits particuliers qui ont engendré les sous-catégories de niveau inférieur : groupe turbopropulseur ; groupe turboréacteur ; turbomoteur. Par conséquent, tous les embranchements d'un même descripteur doivent avoir un certain nombre de sèmes communs, de gènes communs vu qu'ils découlent, ou qu'ils descendent d'une même notion. Et ils doivent, en même temps, être mutuellement exclusifs, en lecture horizontale.

Outre les relations génériques, les cases des arbres de domaine terminologiques, en lecture verticale, entretiennent entre elles des relations partitives. La relation partitive est courante dans les nomenclatures. Elle est très voisine de la relation générique. Si nous avons eu, comme descripteur, le mot « tuyère d'éjection », toutes les parties de cet ensemble auraient, avec cette entité, une relation partitive. Rallonge, buse d'éjection sont, en effet, des parties de la tuyère d'éjection d'un moteur d'avion à réaction. Autre

exemple plus simple : four, porte de four, voûte de four. Les deux derniers syntagmes sont des parties de l'entité « four ». La relation partitive est la division d'un tout en des parties plus réduites. La relation générique est la spécialisation d'une particularité intrinsèque à une notion et donnant une sous-notion à la fois distincte et plus fine.

La terminologie permet, par l'établissement d'équivalences, de jeter un pont entre les cultures. C'est la terminologie bilingue ou multilingue. Un arbre de domaine traitant d'un sujet dans une langue peut-il être traduit dans une autre langue par la simple substitution de mots en présence par des équivalents en langue d'arrivée ? Ce n'est pas toujours le cas, car le découpage des réalités ne s'effectue pas de façon symétrique dans des cultures différentes. L'établissement d'un arbre bilingue ou multilingue pose les mêmes problèmes que la terminologie multilingue. Plusieurs situations peuvent se présenter.

Cas 1 : Équivalence exacte

Tout terminologue, tout traducteur rêve d'une telle situation : les mots d'une langue à l'autre se recoupent intégralement. Exemples :

inlet duct = prise d'air, manche d'entrée d'air.

exhaust nozzle = buse d'éjection.

Cas 2 : Équivalence partielle

La langue d'arrivée fournit un terme dont la signification est partiellement équivalente au terme de la langue de départ. Exemples :

exhaust duct = canal d'éjection

control = commande

management = gestion

Cas 3 : Non-équivalence

Cette situation peut revêtir plusieurs formes :

1. Absence d'une dénomination dans l'une des deux langues. Exemple :

Ø rallonge

2. Présence d'un terme au sens large dans une langue avec plusieurs équivalents possibles dans l'autre. Exemple : *universal joint*

1. joint universel (gén.)

2. joint cardan

3. joint homocinétique.

On résout ces problèmes comme en terminologie. Dans le cas 2, on peut utiliser le descripteur disponible mais en se servant de sous-catégories pour en préciser le sens. Si, dans un arbre, nous avons « *exhaust duct* » qui engendrerait « *exhaust nozzle* », nous déduirions que le terme « *exhaust duct* » serait employé dans le sens de « canal d'éjection ». De même, on pourrait, pour traduire « rallonge » qui n'a pas d'équivalent anglais, proposer « *exhaust duct* » mais en le mettant sur le même plan (horizontalement, cela s'entend) que « *exhaust nozzle* ». On devinerait alors qu'il s'agit du canal dont la buse d'éjection constitue la partie terminale.

Une autre solution consisterait à recourir à l'emprunt intégral. La langue anglaise ne distingue pas la rallonge du canal d'éjection ; nous l'appellerions « rallonge » en anglais. Mais en avons-nous le droit ? On pourrait aussi recourir à la néologie au risque de générer un barbarisme que l'usage boudera (hambourgeois) ou user d'une circonlocution (*scramjet* : statoréacteur supersonique). Ce dernier exemple révèle les limites de l'arbre de domaine. Jusqu'à quel degré de finesse faut-il aller ? L'anglais éprouve le besoin de créer un mot nouveau, « *scramjet* », mot valise résultant du télescopage des éléments du syntagme « *supersonic combustion ramjet* ». Le français, pour le suivre sur ce sentier glissant, propose « statoréacteur supersonique », terme aussi redondant que l'anglais : l'effet statoréacteur (*ram effect*) n'entre en jeu qu'aux vitesses transsoniques et su-

personiques. Dernière solution... de facilité : nier le mot problème en ne le faisant pas figurer dans l'arbre. Mais notre arbre serait-il tout aussi valide ? Comment s'en assurer.

Un arbre de domaine peut, en effet, présenter des faiblesses qui nuisent à sa cohérence. Deux types de lacunes peuvent se présenter. La première s'applique à l'arbre bilingue. Les recoupements entre deux termes homologues dans les deux langues sont insuffisants. La seconde faiblesse se rapporte à l'arbre unilingue : il peut ne pas être cohérent ou comporter des erreurs.

Les erreurs, en plan horizontal, découlent de la position au même niveau de mots n'ayant pas la même extension. Il pourrait, par exemple, s'agir d'un générique à côté d'un spécifique. Dans un tel cas, il suffirait de recourir à un dictionnaire afin de trouver un mot ayant la même extension que les autres mots de même niveau. Parfois, il peut s'agir d'une simple erreur de dénomination : on n'a pas le mot exact. Alors, faute de mieux, on use d'un mot au sens vague.

Les erreurs au plan vertical sont plus difficiles à corriger. Elles sont souvent dues à de mauvaises inclusions. Selon l'arbre de la figure 7, le « groupe motopropulseur » appartient à la catégorie des « turbomachines aéronautiques ». Or il s'agit de deux technologies fondamentalement opposées. Pour corriger ce type d'erreur, il faut, par abstraction, trouver une notion qui inclurait ces deux catégories tout en respectant leur particularité. Ce serait « moteurs d'avions » qui pourrait se diviser en deux sous-catégories « moteurs à piston » et « turbomachines ».

Afin de repérer les erreurs, il convient de soumettre l'arbre à un test de validation. Le premier est le test de la définition. Il consiste à prendre un descripteur quelconque et d'en vérifier la définition conjointement dans un dictionnaire technique ou, mieux, dans une encyclopédie, et par rapport à l'arbre.

Exemple : Selon la figure 6, le statoréacteur n'a pas d'éléments tournants, appartient au groupe turboréacteur, lequel est une turbine à gaz utilisé comme moteur d'avion. Ce qui donne la relation suivante :

Statoréacteur ← sans éléments tournants ← groupe turboréacteur
 turbine à gaz ← moteurs d'avion.

Définition de l'encyclopédie :

Statoréacteur : « Le *statoréacteur* ou *tuyère thermopropulsive* représente la forme la plus élémentaire du *moteur à réaction*. Il consiste essentiellement en une chambre de combustion ouverte à ses deux extrémités (entrée et sortie). Il ne comporte *aucune pièce en mouvement*... Sa poussée, nulle à vitesse nulle, augmente avec la vitesse de translation, de sorte que cet engin pourrait devenir spécialement intéressant pour la *propulsion des avions* à des vitesses soniques et surtout supersoniques. »

Extrait de l'*Encyclopédie pratique Quillet : applications mécaniques* (1957), p. 1221.

Cet extrait recouvre à peu près les embranchements de notre arbre. Comment ?

moteurs d'avion → propulsion des avions
 turbine à gaz → Ø
 groupe turboréacteur → moteur à réaction
 sans éléments tournants → aucune pièce en mouvement
 Statoréacteur → statoréacteur ; tuyère thermopropulsive

L'application de ce test en deux ou trois points de l'arbre peut suffire pour en établir la cohérence.

Le second test de validation, complémentaire du premier, consiste à soumettre l'arbre à l'analyse d'un spécialiste du domaine. On pourrait ainsi découvrir que pour un sujet donné, il peut exister mille et une façons de l'organiser sous forme d'arbre, qu'un

arbre de domaine demeure une intervention, une tentative d'organisation plus ou moins arbitraire d'un domaine et, qu'enfin, les liaisons que l'on tisse entre les descripteurs ne valent que par rapport à l'objectif que l'on s'est assigné.

Quoique laborieux à dresser, l'arbre, une fois dessiné, présente d'énormes avantages. Il permet, d'un seul coup d'œil, d'embrasser tout un champ lexical, de visualiser les relations entre les différentes notions. Son élaboration oblige à un travail de lecture, afin de dénombrer toutes les notions pertinentes, et d'abstraction, afin de les intégrer dans un tout cohérent.

Cependant, le souci de clarification qui pousse le terminologue à établir un arbre peut rendre le produit final inintelligible. Un arbre aussi géant que le *Thesaurus of Terms on Copper Terminology* (établi par l'UNESCO) qui mesure 1,50 m, perd toute clarté puisqu'il nécessite un long moment de recherche pour y déceler un terme d'intérêt et un temps équivalent pour repérer toutes les ramifications pertinentes entre ce terme et les autres. Jusqu'où ne pas aller trop loin ?

MASSIVA N. ZAFIO

BIBLIOGRAPHIE

- AUGER, Pierre et Louis-Jean ROUSSEAU (1978) : *Études, recherches et documentation, Méthodologie de la recherche terminologique*, Québec, OLF.
- DUBUC, Robert (1978) : *Manuel pratique de terminologie*, Montréal, Linguattech.
- HÉZELLES, Norbert et Nadine VIGIER (1979) : *la Généalogie, manuel de l'amateur*, Paris, Hachette.
- LAURIELLE, Marie-Thérèse (1981) : *le Thésaurus, son rôle, sa structure, son élaboration*, Villeurbanne, Presse de l'ENSP.
- AFNOR (1973) : « Norme expérimentale, Documentation », *Règles d'établissement des thésaurus en langue française* (Z 47-100).
- AFNOR (1980) : « Norme expérimentale, Documentation », *Principes directeurs pour l'établissement des thésaurus multilingues* (Z 47-101).