



TOTh 09

Terminologie & Ontologie : Théories et Applications

Actes de la troisième conférence TOTh - Annecy - 4 & 5 juin 2009



Institut Porphyre
Savoir et Connaissance

Dans un monde où la communication et le partage d'information sont au cœur de nos activités, les besoins en terminologie se font de plus en plus pressants. Il est devenu impératif d'identifier les termes employés et de les définir de façon consensuelle et cohérente tout en préservant la diversité langagière.

La terminologie, en tant que discipline scientifique, se fonde sur une conceptualisation d'un domaine et sur les mots pour en parler. Elle se doit donc de concilier un point de vue linguistique et un point de vue ontologique. Elle doit également, dans une société numérique où les connaissances constituent la principale richesse, pouvoir être opérationnalisée à des fins de traitement de l'information.

Les conférences TOTh se situent dans le prolongement des colloques annuels de la Société française de terminologie organisés en décembre à Paris (Ecole normale supérieure de la rue d'Ulm). Planifiées à mi-parcours, au mois de juin à Annecy (Polytech'Savoie), elles en complètent l'offre et proposent des conférences avec appel à communications, comité de lecture et publication des actes.

Les conférences TOTh ont pour objectif de rassembler industriels, chercheurs, utilisateurs et formateurs dont les préoccupations relèvent à la fois de la terminologie et de l'ontologie et, de façon plus générale, de la langue et de l'ingénierie des connaissances. Elles se veulent un lieu d'échange et de partage où sont exposés problèmes, solutions et retours d'expériences tant sur le plan théorique qu'applicatif ; ainsi que les nouvelles tendances et perspectives des disciplines associées : terminologie, langues de spécialité, linguistique, intelligence artificielle, systèmes d'information, ingénierie collaborative, etc.

Christophe Roche, Président du Comité Scientifique

<http://www.porphyre.org>



Institut Porphyre
Savoir et Connaissance

ISBN 978-2-9536168-0-4
EAN 9782953616804

Publications précédentes

TOTh 2007

Actes de la première conférence TOTh - Annecy - 1^{er} juin 2007

TOTh 2008

Actes de la deuxième conférence TOTh - Annecy - 5 et 6 juin 2008

Commandes à adresser à : toth@porphyre.org

Titre : TOTh 2009. *Actes de la troisième conférence TOTh - Annecy - 4 & 5 juin 2009*

Editeur : Institut Porphyre, *Savoir et Connaissance*

<http://www.porphyre.org>

Annecy, 2009

ISBN 978-2-9536168-0-4

EAN 9782953616804

© Institut Porphyre, *Savoir et Connaissance*



Actes de la conférence

TOTh 2009

Annecy – 4 & 5 juin 2009

avec le soutien de :

- Société française de terminologie
- Association Européenne de Terminologie
- Ecole d'ingénieurs Polytech'Savoie – Université de Savoie
- Université de Sorbonne nouvelle
- Association EGC (Extraction et Gestion des Connaissances)
- ISKO (International Society for Knowledge Organization) France



Institut Porphyre
Savoir et Connaissance

<http://www.porphyre.org>

Comité scientifique

Président du Comité Scientifique : Christophe Roche

Comité de pilotage

Loïc Depecker	Professeur, Université de Sorbonne nouvelle
André Manificat	Directeur, GRETh
Christophe Roche	Professeur, Université de Savoie
Philippe Thoiron	Professeur émérite, Université de Lyon II

Comité de programme

Bruno de Bessé	Professeur, Université de Genève
Pierre Blanc	EDF SEPTEN
Danièle Bourcier	CNRS, CERSA Paris
Marc van Campenhoudt	Professeur, Termisti, ISTI, Bruxelles
Danielle Candèl	CNRS, Université Paris Diderot
Stéphane Chaudiron	Professeur, Université de Lille III
Viviane Cohen	France Télécom, Paris
Rute Costa	Professeur, Université Nouvelle de Lisbonne
Luc Damas	MCF, Université de Savoie
Sylvie Desprès	MCF, Université Paris XIII
François Gaudin	Professeur, Université de Rouen
Anne-Marie Gendron	Chancellerie fédérale suisse, Section de terminologie
Jean-Yves Gresser	ancien Directeur à la Banque de France
Ollivier Haemmerlé	Professeur, Université de Toulouse
Jean-Paul Haton	Professeur, Université de Nancy 1
Michèle Hudon	Professeur, Université de Montréal
John Humbley	Professeur, Université Paris 7
Michel Ida	Directeur MINATEC, CEA
Hendrik Kockaert	Professeur, Lessius Hogeschool (Anvers)
Michel Léonard	Professeur, Université de Genève
Pierre Lerat	Professeur honoraire, Université Paris XIII
Widad Mustafa	Professeur, Université de Lille III
Henrik Nilsson	Terminologocentrum TNC, Suède
Jean Quirion	Professeur, Université du Québec en Outaouais
Renato Reinau	Suva, Lucerne
François Rousselot	MCF, Université de Strasbourg
Gérard Sabah	CNRS, Orsay
Michel Simonet	CNRS Grenoble
Marcus Spies	Professeur, Université de Munich
Dardo de Vecchi	Professeur associé, Euromed-Management

Comité d'organisation :

Responsable : Luc Damas
Samia Chouder, Joëlle Pellet

Avant propos



Dès la troisième édition, les conférences TOTh ont trouvé une structuration qui traduit bien à la fois le caractère scientifique et pluridisciplinaire de la terminologie et l'intérêt de notre communauté pour d'autres domaines partageant des préoccupations communes.

Ainsi, la conférence d'ouverture a été donnée par une personnalité invitée issue d'une discipline différente de la nôtre – ici la phylogénèse – mais pour laquelle le langage et la pensée jouent également un rôle primordial.

Les contributions se sont réparties naturellement, et par le jeu des évaluations de façon équitable, en trois groupes ayant donné lieu à trois sessions.

Le premier groupe a rassemblé les articles portant principalement sur la dimension linguistique de la terminologie. Ont été abordés l'extraction terminologique à partir de dictionnaire, la place accordée aux corpus dans la construction de terminologies, l'acquisition de connaissances à partir de textes et l'apport des ressources linguistiques issues du web.

La deuxième session s'est donc logiquement intéressée à la dimension conceptuelle de la terminologie. Les notions de concept, de relation, d'ontologie ont été au cœur des présentations portant sur les cartes conceptuelles pour les bibliothèques numériques, les relations dynamiques et les graphes conceptuels, l'alignement d'ontologies et l'accès multilingue aux ontologies.

Enfin, la troisième session a été consacrée à la présentation de plusieurs applications terminologiques pour des secteurs aussi différents que l'ingénierie nucléaire, l'informatique, le domaine bancaire ou l'agriculture biologique. Il est à souligner que ces applications ont permis d'aborder différents points théoriques tels que la variation terminologique, la diachronie ou la structure des dictionnaires.

La richesse des débats qui ont animé ces deux jours de conférence – chaque présentation, questions comprises, s'est vue allouer plus de quarante cinq minutes de temps de parole – a été certainement une des plus belles récompenses pour les participants de TOTh 2009.

Christophe Roche

Président du Comité Scientifique

Table des matières

CONFERENCE INVITEE

<i>La nomenclature biologique aujourd'hui : que reste-t-il de Linné ?</i>	1
Michel Laurin	

SESSION 1

<i>Approche lexico-sémantique de l'extraction terminologique : utilisation de ressources lexicographiques et validation sur corpus</i>	19
Bertrand Gaiffe, Evelyne Jacquey, Laurence Kister	
<i>Quelle place accorder aux corpus dans la construction d'une terminologie ?</i>	33
Marie Calberg-Challot, Pierre Lerat, Christophe Roche	
<i>Extraction de connaissances orientées évolution dans les textes techniques</i>	53
Kata Gabor, François Rousselot, François De Bertrand de Beuvron	
<i>Corpus et Web : deux alliés pour la construction de l'enrichissement automatique de classes conceptuelles</i>	73
Nicolas Béchet, Mathieu Roche, Jacques Chauché	

SESSION 2

<i>Following the path between conceptual maps and visual thesauri</i>	93
Olga Bessa Mendes	
<i>Dynamic concept relations: a definition and representation proposal</i>	107
Chiara Messina	
<i>Construction et alignement d'ontologies pour évaluer le risque alimentaire</i>	127
Liliana Ibanescu, Patrice Buche, Juliette Dibie-Barthélemy	
<i>Accès multilingue à une ontologie par des correspondances avec un lexique pivot</i>	143
David Rouquet, Hong-Thai Nguyen	
<i>La reformulation : processus dynamique d'acquisition des connaissances. Le cas du discours technique arabe d'Internet</i>	161
Andrée Affeich	

SESSION 3

<i>Structuration d'un dictionnaire de spécialité pour sa publication sur internet. Bénéfices du langage XML</i>	181
Jacques Joseph	
<i>Mémoire du Club informatique des grandes entreprises françaises (CIGREF) : nouveau plan de classement</i>	197
Jean-Yves Gresser, M.P. Lacroix	
<i>Les secteurs d'activité à l'épreuve du discours</i>	217
Frédéric Erlos	
<i>De l'agriculture biologique aux espaces naturels : une étude des syntagmes terminologiques à l'intérieur des textes de spécialité</i>	235
Elisa Lavagnino	
<i>Pages blanches</i>	253

La nomenclature biologique aujourd'hui : que reste-t-il de Linné ?

Michel Laurin

Résumé : La nomenclature biologique est une activité essentielle étant donné que nous connaissons plus d'un million et demi d'espèces d'êtres vivants. La nomenclature Linnéenne-Stricklandienne, utilisée par la plupart des systématiciens depuis deux siècles, repose sur l'utilisation d'un seul type et d'un rang absolu (catégorie Linnéenne). Or, si le type a bien une existence objective, les catégories Linnéennes en sont dépourvues. De plus l'utilisation d'un seul type et d'un rang ne peut aucunement fournir une délimitation unique et stable du contenu des taxons. Cette absence de délimitation, apparemment recherchée par les commissions ayant rédigé les codes formalisant cette nomenclature, semble inappropriée puisque généralement, les scientifiques tentent de définir les termes techniques aussi précisément que possible afin de communiquer efficacement. Ceci est démontré par des comparaisons avec la géochronologie, la géopolitique et la chimie. Tout ceci explique le développement, à partir des années 1980, d'une nomenclature phylogénétique reposant sur l'utilisation d'au moins deux déterminants (types) et permettant de délimiter précisément les taxons à l'aide d'une phylogénie (arbre évolutif).

Mots-clés : Terminologie, taxonomie, systématique, codes de nomenclature

1. Introduction

La nomenclature biologique est une activité essentielle étant donné que nous connaissons plus d'un million et demi d'espèces d'êtres vivants. Il n'y a des spécialistes de ce domaine que depuis le 18^{ème} siècle environ, Linné ayant été un des premiers auteurs à consacrer la plus grande partie de son activité à ce domaine. Cependant, divers auteurs du anciens s'y sont adonnés, dont Aristote est sans doute le plus connu et l'un des plus anciens (Voultsiadou & Vafidis 2007) et on retrouve des traces de cette activité dès la Genèse, qui fut écrite il y a presque 3000 ans. Par exemple, on peut y lire (Genèse 2 : 19-20) : "Le Seigneur Dieu modela du sol toute bête des champs et tout oiseau du ciel qu'il amena à l'homme pour voir comment il les désignerait". Il est donc clair que le besoin de classer la diversité naturelle s'est fait sentir très tôt dans l'histoire de l'humanité.

La nomenclature biologique comporte diverses particularités liées à la nature des êtres vivants et surtout à l'évolution biologique. En effet, la plupart des objets que nous classons forment justement des classes (d'où le verbe "classer", soit "ranger dans des classes"). Les entités formant des classes sont identiques, ils peuvent exister n'importe où dans l'univers, et ne sont pas temporellement délimités. Par exemple, un atome d'hydrogène ou d'hélium d'un isotope donné ne diffère aucunement des autres atomes du même élément et du même isotope. De tels atomes peuvent exister n'importe où dans l'univers. Ils peuvent avoir existé depuis environ 300 000 ans après le Big Bang (temps nécessaire pour que la matière refroidisse suffisamment pour que les atomes capturent leurs électrons) et peuvent exister indéfiniment, si l'univers est ouvert. Les êtres vivants, au contraire, sont spatio-temporellement délimités. Ainsi, les chats (*Felis domesticus*) et les humains (*Homo sapiens*) n'existent que sur terre, n'ont jamais existé où que ce soit ailleurs, et n'existeront jamais sur une autre planète sauf si nous les y transportons dans un futur lointain. Ils n'existent que depuis moins d'un million d'années et n'existeront pas éternellement. De plus, il n'y a pas deux chats (ou d'êtres humains) identiques. Même les jumeaux dits identiques ne le sont que de par leur génome, mais des différences apparaissent dans leur personnalité, leur système immunologique, etc. De tels êtres ne forment pas des classes, mais bien des individus, selon de nombreux auteurs (Ereshefsky 2007), même si certains préfèrent y voir un type d'entité intermédiaire entre des universaux formant des classes et des individus (Rieppel 2005). Depuis quelques décennies, on considère souvent que les taxons ne forment pas des classes, mais plutôt des individus, ce qui explique la difficulté de les définir d'après leurs propriétés intrinsèques.

On comprendra donc qu'en nomenclature biologique, l'étymologie s'explique, mais n'explique pas. En effet, les taxons sont souvent nommés d'après une propriété partagée par la plupart de ses membres mais cette propriété n'est ni nécessaire, ni suffisante pour déterminer l'appartenance d'un organisme à un taxon. Ainsi, le taxon Tetrapoda inclut des animaux ayant quatre membres, mais certains animaux dépourvus de membres, comme les serpents et les gymnophiones, en font aussi partie. Ils descendent d'ancêtres qui possédaient quatre membres et sont ainsi unanimement classés parmi les tétrapodes. De plus, si des animaux acquéraient quatre membres de façon convergente, ils ne seraient pas des tétrapodes. On voit ainsi qu'on ne peut généralement pas utiliser les attributs intrinsèques des organismes directement pour les attribuer à des taxons ; cette attribution dépend en fait plutôt de la position des organismes dans l'arbre évolutif du vivant. Dans ce fonctionnement même, on voit que la classification des êtres vivants opère de façon plus similaire à la généalogie, qui ne concerne que des individus et repose sur l'histoire, qu'à la classification d'universels, qui repose sur leurs propriétés intrinsèques.

L'arbre évolutif du vivant s'impose donc comme la base de toute classification du vivant. Mais alors, comment subdiviser cet arbre, qui forme un tout de sa racine au sommet de chaque branche ? Pour les systématiciens qui continuent à conceptualiser les taxons comme des classes, comment peuvent-ils définir et délimiter ces classes ? Si les taxons sont bien des individus, comment les définir et les délimiter ? Ces questions sont au cœur des débats actuels en nomenclature biologique, et nous verrons que leurs réponses diffèrent profondément entre la nomenclature Linnéenne-Stricklandienne et la nomenclature phylogénétique.

2. Caractéristiques principales de la nomenclature Linnéenne-Stricklandienne

En nomenclature Linnéenne-Stricklandienne, on attribue à chaque taxon un rang absolu ou catégorie Linnéenne. Les catégories principales vont, des plus grandes aux plus petites, du règne à l'espèce. Ainsi, la classification de notre espèce peut être représentée ainsi :

Règne : Metazoa

Embranchement : Chordata

Classe : Mammalia

Ordre : Primates

Famille : Hominidae

Genre : *Homo*

Espèce : *Homo sapiens*

On utilise en plus des préfixes pour augmenter le nombre de rangs disponibles. On peut ainsi former les noms de la superfamille Hominoidea et de la sous-famille Homininae. Notez que pour la série-famille, les terminaisons changent également ; pour les animaux, ces noms se terminent par "-oidea" pour les superfamilles, "-idae" pour les familles, et "-inae" pour les sous-familles.

Chaque taxon est ancré dans la réalité par un type, qui est soit un organisme préservé (squelette, spécimen préservé dans l'alcool, etc.), pour les espèces, soit une espèce, pour les genres, soit un genre, pour les familles. Ainsi, seules les espèces sont directement ancrées à la réalité par les spécimens. Au-dessus de ce rang, le lien est indirect, mais ultimement, même une famille est ancrée dans la réalité car elle est définie par un genre-type, qui est lui-même défini par une espèce-type, qui est elle-même définie par un spécimen-type. En zoologie, les taxons d'un rang supérieur à celui de la série-famille n'ont pas de types ; leur sens est donc un peu moins bien déterminé. Linné n'utilisait pas de types ; ils ont été introduits en nomenclature biologique par le code de Strickland (Strickland *et al.* 1842, 1843), d'où l'expression de nomenclature "Linnéenne-Stricklandienne" proposée par Dubois (2006).

Le principe de priorité, reconnu à divers degrés par tous les codes de nomenclature, stipule que généralement, le premier nom proposé pour un taxon est valide (ainsi que sa définition). Les autres sont soit des synonymes (d'autres noms pour le même taxon), soit des homonymes (le même nom défini autrement).

La définition des noms de taxons dans ce système consiste uniquement en un type et un rang (Laurin 2008a). On caractérise également les taxons par des diagnoses, qui sont des listes de caractères qui sont censées permettre de déterminer si un organisme appartient à un taxon. Ainsi, une diagnose des tétrapodes mentionnera la présence de quatre membres pourvus de doigts, l'absence de branchies, la présence de poumons, etc. Cependant, les diagnoses ne font pas partie des définitions car elles peuvent être modifiées et deux taxons ne peuvent pas être déclarés synonymes parce qu'ils partagent la même diagnose.

La nomenclature Linnéenne-Stricklandienne est régulée par trois codes principaux : le code zoologique pour les animaux, le code botanique pour les plantes et le code bactériologique pour les eubactéries et archées (Laurin 2005). Les limites entre les juridictions de ces codes sont parfois floues et artificielles. Ainsi, certaines bactéries (les cyanobactéries) sont photosynthétiques et sont souvent considérées comme des plantes ; leurs noms sont généralement régulés par le code botanique, même si elles sont plus étroitement apparentées aux autres bactéries qu'aux plantes vertes. De nombreux organismes eucaryotes ont acquis la capacité de photosynthèse par endosymbiose ; ils ont en quelque sorte incorporé des cyanobactéries dans leurs cellules, et ces bactéries sont devenues les chloroplastes. Certaines espèces de ces taxons ont ensuite perdu les chloroplastes. Traditionnellement, les noms de tous les organismes photosynthétiques et des champignons sont régis par le code botanique, alors que les noms de tous les autres eucaryotes sont régis par le code zoologique. Ceci signifie que dans de nombreux groupes d'organismes, les noms de certaines espèces sont régulés par le code zoologique, alors que d'autres espèces étroitement apparentées sont régulées par le code botanique. Pire, dans certains cas, une espèce donnée possède un nom régi par le code botanique, et un autre par le code zoologique !

3. Problèmes causés par la nomenclature Linnéenne-Stricklandienne

Les catégories attribuées aux taxons sont subjectives, car il n'y a ni famille, ni embranchement dans la nature. Ceci signifie que les systématiciens peuvent altérer ces rangs, ce qui résulte souvent en un changement de la délimitation du taxon. En effet, l'utilisation d'un seul type et d'un rang subjectif ne peut pas délimiter un taxon (Figure 1).

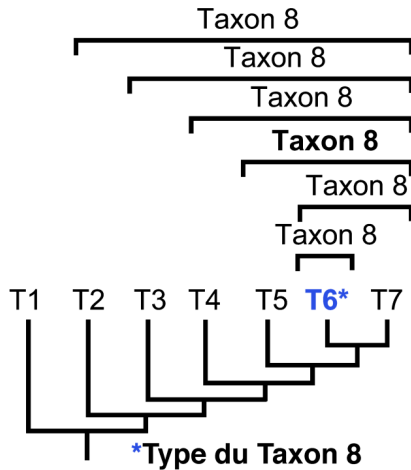


Figure 1. Délimitation floue en nomenclature Linnéenne-Stricklandienne

Par exemple, dans la Figure 1, si T1–T7 sont des genres et que T8 est une famille, la désignation de T6 comme type ne permet pas de délimiter T8. Plusieurs délimitations possibles sont indiquées sur la figure. Même si l’auteur ayant érigé T8 a spécifié qu’il souhaitait y inclure T5–T7 (délimitation en caractères gras), cette délimitation n’est pas contraignante ; les autres systématiciens ne sont pas tenus de la respecter, même s’ils n’ont aucune raison objective de redélimiter T8.

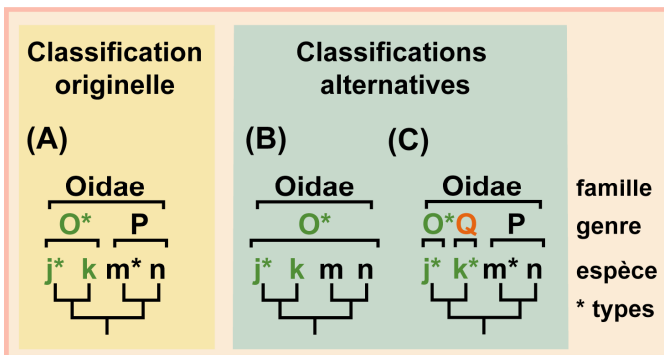


Figure 2. Synonymie et instabilité en nomenclature Linnéenne-Stricklandienne.
Adapté de Laurin (2008b)

Cette délimitation floue est en partie liée à des synonymies potentielles de taxons étroitement apparentés, synonymie liée à la nature subjective des catégories Linnéennes (Lee et Skinner 2007). Ceci peut être illustré par un

exemple (Figure 2). Supposons qu'on découvre quatre nouvelles espèces (j, k, m et n) étroitement apparentées et qu'une phylogénie indique que ces espèces forment deux clades (Figure 2A ; le terme "clade" désigne un groupe d'organismes biologiques comprenant un ancêtre et tous ses descendants). Le systématique qui décrira ces espèces pourrait alors choisir de reconnaître deux genres (O et P) et une nouvelle famille (Oidea). Les espèces j et m pourraient être désignées comme types des genres O et P et le genre O, comme type de Oidea. Sans aucune raison objective, un autre taxonomiste pourrait décider de mettre le genre P en synonymie avec O, étendant ainsi le contenu de O (Figure 2B). Un troisième taxonomiste pourrait décider, au contraire, de reconnaître un genre additionnel (Q) en choisissant l'espèce k comme type (Figure 2C), ce qui a pour résultat de réduire le contenu du genre O à sa seule espèce-type (j). D'autres considérations mènent également à considérer les catégories Linnéennes comme de dangereux reliquats d'un système de nomenclature périmé (Minelli 2000).

Puisque les codes Linnéens-Stricklandiens visent à ne pas délimiter les taxons précisément, toutes ces alternatives (et bien d'autres encore) sont simultanément et indéfiniment valables, ce qui signifie que le sens des noms est ambigu. Par exemple, le genre O peut contenir une seule espèce (j), deux (j, k), ou quatre (j, k, m et n), pour ne mentionner que ces trois possibilités. Ceci n'est pas qu'un problème purement théorique. Il affecte de nombreux taxons (peut-être la plupart) depuis avant même l'entrée en vigueur des codes de nomenclature qui ont justement été inaugurés pour réduire le chaos nomenclatural qui prévalait déjà dans les années 1840 (Strickland *et al.* 1842). Il semble malheureusement que les codes Linnéens-Stricklandiens n'ont eu, de ce point de vue, qu'un succès mitigé (Laurin 2008b). Ainsi, Rowe et Gauthier (1992) ont inventorié pas moins de 10 sens du nom Mammalia (taxon qui inclut les mammifères) dans la littérature scientifique publiée dans des années 1960 à 1990. On pourrait argumenter que la classe Mammalia n'a pas de type et que le code zoologique ne peut donc pas stabiliser son sens, mais de tels problèmes affectent tout autant les taxons des séries-famille, genre et espèce, qui ont pourtant des types. Ainsi, T. M. Keesey a récemment répertorié six sens du nom Hominidae (Laurin & Bryant 2009), et un débat a actuellement lieu concernant la classification des très nombreuses espèces (plus de 1000) du genre *Rana* (Frost *et al.* 2006 ; Hillis 2007).

4. Comparaisons avec d'autres domaines

La nomenclature Linnéenne-Stricklandienne semble relativement isolée des autres domaines de la connaissance humaine en tentant de ne pas définir précisément ses termes techniques. En fait, on pourrait argumenter que les noms de taxons tels qu'ils sont définis dans ce système ne sont pas véritablement des termes car ils ne correspondent pas à une délimitation stable de la réalité (Calberg-Challot *et al.* ce tome). Dans les autres domaines, on cherche généralement à préciser autant que possible les termes techniques.

Ainsi, la géochronologie utilisait autrefois des sections-types analogues aux types de la nomenclature Linnéenne-Stricklandienne. Ces sections-types définissaient des périodes plus ou moins longues, mais elles ne les délimitaient pas. Ainsi, la section-type du Lutétien ne représente qu'une petite proportion du temps représenté par ce dernier (Figure 3). Pendant les vingt dernières années, les géologues ont remplacé ces sections-types par des limites-types, appelées GSSPs, pour "Global Stratotype Section and Point" (Gradstein *et al.* 2004 : 20–21). Ces GSSPs se présentent sous forme d'un marqueur précis (par exemple, un clou enfoncé dans une section) identifiant précisément la limite inférieure d'une division temporelle (étage, période ou ère). La limite inférieure de la division suivante (plus récente) sert simultanément de limite supérieure à la division précédente. Ainsi, on est passé d'un système de nomenclature flou à un système très précis.

Les chimistes et les physiciens utilisent depuis longtemps une classification très précise de la matière fondée sur le nombre de protons dans le noyau atomique. Les éléments sont ensuite regroupés par familles chimiques en fonction du nombre d'électrons de valence, ce qui donne le fameux tableau des éléments initialement proposé par le Chimiste Russe Mendeleev en 1869. Ainsi, le lithium (Li) comporte trois protons et un électron de valence. Le sodium (Na) et le potassium (K) possèdent 11 et 19 protons et un seul électron de valence, ce qui explique qu'on les réunisse dans la même famille que le lithium. Les adeptes de la nomenclature Linnéenne-Stricklandienne devraient peut-être tenter de convaincre les chimistes et les physiciens d'un système de nomenclature dans lequel "lithium" désigne tantôt les atomes comportant 3 protons, tantôt ceux comportant 3 ou 11 protons, et tantôt ceux comportant 3, 11, ou 19 protons et dans lequel tous ces sens sont simultanément et indéfiniment valides. Un consensus sur le sens de ce terme devrait alors émerger (ou pas ?) par un accord spontané (et peut-être temporaire).

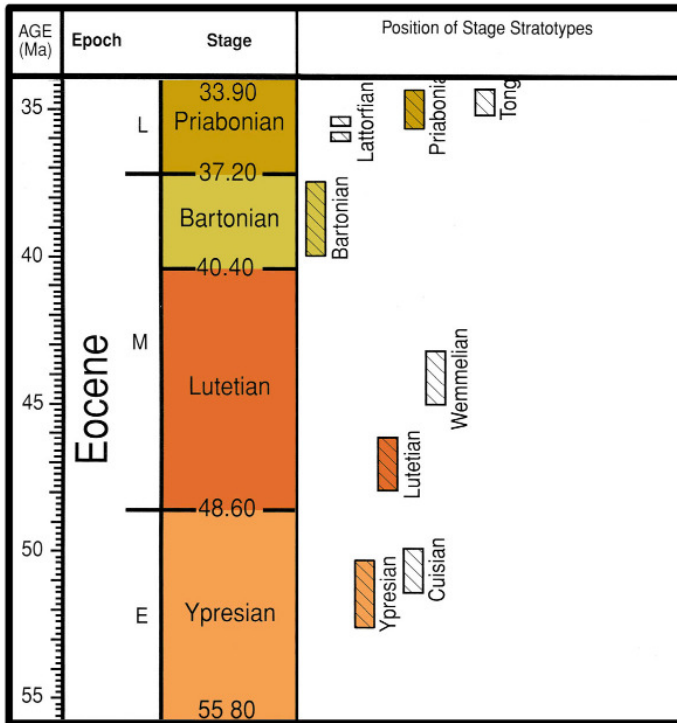


Figure 3. Nomenclature géochronologique. Adapté de Gradstein *et al.* (2004)

Les entités géographiques, comme les taxons, possèdent des rangs (continent, pays, province ou état, ville, quartier, etc.). On peut même considérer que les entités territoriales ont des types (chefs-lieux, capitales). Mais quel géographe ou politicien se contenterait de définir les entités territoriales par un rang et un chef-lieu sans aussi spécifier des frontières ou limites ? Imaginez la carte de l'Europe si la France était définie comme l'entité territoriale de rang Pays ayant Paris pour Capitale, et si tous les autres pays étaient définis de manière similaire. Évidemment, les géographes et les politiciens délimitent les entités territoriales depuis longtemps de façon précise en utilisant souvent soit des frontières naturelles (chaînes de montagnes, fleuves, côtes, etc.), soit des bornes placées aux frontières, comme c'était le cas dans la Mésopotamie antique, région plate se prêtant peu à l'utilisation de frontières naturelles. Par analogie avec les taxons, on peut considérer que les pays peuvent être mis en synonymie (conquis) ou redélimités (lors de conflits territoriaux). Cependant, contrairement aux taxons, les frontières des pays sont âprement défendues depuis des millénaires, comme en témoigne la stèle des Vautours, conservée au Louvre et datant du 2450 AC (Bottéro 1994).

Ceci est logique, puisque le territoire est précieux. Devrait-on également conclure, toujours par analogie, que les taxons, qui représentent le produit du travail des taxonomistes, ne sont pas précieux ? Les taxonomistes n'attachent-ils donc aucune valeur à leur propre travail ?

5. Pourquoi cette nomenclature est-elle toujours utilisée ?

Cette démonstration par l'absurde de l'isolement de la nomenclature Linnéenne-Stricklandienne conduit naturellement à se questionner sur la persistance d'un tel système dans la biologie du 21^{ème} siècle. Bien qu'il soit difficile d'apporter une réponse définitive, deux facteurs ont probablement joué un rôle important. Le premier est de nature historique. Lorsque le premier code de nomenclature fut formulé (Strickland *et al.* 1842), la phylogénétique commençait à peine et la commission qui formula le code pensa sans doute qu'il serait utopique de fonder un code de nomenclature sur une phylogénie qui était alors largement inconnue et donc l'existence même n'était pas acceptée par tous les naturalistes. Même Darwin, qui faisait partie de cette commission, fut sans doute de cet avis car il fut contraint d'accepter des groupes paraphylétiques (n'incluant pas tous les descendants d'un ancêtre) dans ses fameuses monographies sur les cirripèdes, faute de pouvoir proposer une phylogénie détaillée (Padian 1999). Déjà à cette époque, la nomenclature biologique avait une longue histoire que le code de Strickland respecta. Ainsi, les codes de nomenclature Linnéens-Stricklandiens représentent l'aboutissement d'une tradition de plus de deux siècles et demi. Il n'est donc pas surprenant que de nombreux systématiciens ne pensent même pas à remettre en cause un système que des générations de taxonomistes ont utilisé, et qu'ils soient même choqués par l'initiative de certains de remplacer ce système.

La seconde raison de la persistance de la nomenclature Linnéenne-Stricklandienne est simplement l'absence, jusqu'à un passé très récent, d'une alternative viable. En effet, la seule alternative ayant réussi à obtenir le soutien d'une communauté internationale de taxonomistes est la nomenclature phylogénétique, bien que quelques autres aient été proposées (e.g. Dubois 2006 ; Béthoux 2007). Celle-ci fut initialement utilisée il y a un peu plus de 20 ans dans un article sur l'origine des oiseaux (Gauthier 1986). Les principes de base de cette nomenclature furent explicités quelques années plus tard (de Queiroz et Gauthier 1990). Divers autres systématiciens commencèrent alors à utiliser ce système pendant les années 1990 (e.g. Laurin 1991 ; Wolsan 1993 ; Bryant 1994). Une première version du PhyloCode, un code de nomenclature pylogénétique, fut

publiée sur l'Internet en 1998 par un comité de 30 systématiciens, mais une société internationale pour encadrer son développement (ISPN : International Society for Phylogenetic Nomenclature) ne fut fondée qu'en 2004, lors d'un congrès réunissant 70 systématiciens de 11 pays (Laurin et Cantino 2004). Ce code devrait officiellement entrer en vigueur d'ici quelques années, mais il est déjà utilisé par plusieurs de dizaines (si ce n'est de centaines) de systématiciens dans de nombreux pays.

6. Nomenclature phylogénétique

Comment définit-on les noms de taxons en nomenclature phylogénétique ? Il y a trois façons possibles : par nœud (nodale), par branche, ou par apomorphie (Figure 4). Une définition nodale utilise au moins deux types (appelés déterminants, pour les différencier des types Linnéens-Stricklandiens) internes. Elles peuvent prendre la forme : "Le plus petit clade comprenant A et B", ou (ce qui est identique) : "le dernier ancêtre commun de A et B, et tout les descendants de cet ancêtre", où A et B désignent des espèces (ou bien des individus). Le groupe ainsi défini est identifié en rouge sur la figure.

Une définition par apomorphie utilise une apomorphie (nouveau caractère évolutif ou caractère récent) et au moins une espèce (ou un individu) comme déterminants. Elle peut prendre la forme : "Le clade délimité par l'apomorphie M synapomorphique avec A. Le terme "synapomorphique" signifie que l'apomorphie M n'a pas disparue avant que n'apparaisse l'espèce A, ou que cette dernière n'a pas acquis cette apomorphie de façon convergente. Ainsi, la définition "Aves est le clade délimité par l'apparition de l'aile synapomorphique avec *Passer domesticus* (le moineau domestique)" désigne l'ensemble des oiseaux, mais pas les chauve-souris, ptérosaures (reptiles volants du Mésozoïque), ou insectes, qui ont tous acquis des ailes par convergence (le dernier ancêtre commun de tous ces animaux n'avait pas d'ailes). Le groupe ainsi défini est identifié en bleu sur la figure.

Une définition par branche utilise au moins une espèce (ou individu) comme déterminant interne et au moins une espèce comme déterminant externe. Elle peut prendre la forme : "Le plus grand clade incluant A mais pas Z". Le groupe ainsi défini est identifié en vert sur la figure.

Notez que si seules les espèces A, B et Z étaient initialement connues (C et D n'étant découvertes que plus tard), les taxons correspondant à ces trois définitions auraient même contenu, mais comme les définitions sont de trois types, ces taxons ne sont pas synonymes. D'ailleurs, l'éventuelle

découverte d'espèces supplémentaires (C, D) démontrerait bien que ces taxons sont bien distincts. Même dans ce cas, aucune décision subjective ne doit être prise. L'espèce C appartient seulement au clade défini par une branche (en vert), alors que l'espèce D appartient à ce taxon ainsi qu'au taxon fondé sur une apomorphie.

Ce système n'utilise que des définitions comportant des déterminants et des phylogénies de référence pour délimiter les taxons. Les catégories Linnéennes, qui sont artificielles et dépourvues d'existence objective, ne sont pas utilisées. Deux taxons ne sont synonymes que s'ils réfèrent au même clade. Par exemple, si après que les trois définitions mentionnées ci-dessus soient publiées, un taxonomiste définissait un autre taxon comme "le plus grand clade incluant B mais pas Z", ce dernier serait synonyme du taxon fondé sur une branche (en vert) ; le premier taxon nommé et défini pour ce clade aurait priorité.

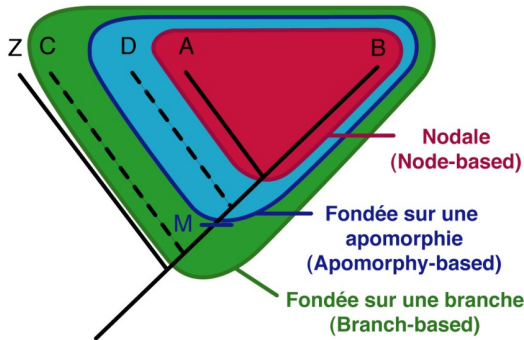


Figure 4. Définitions en nomenclature phylogénétique. Les lettres A–D et Z représentent des espèces ; M représente une apomorphie (caractère récent) Adapté de Laurin (2008a)

Le PhyloCode ne régira pas (au moins initialement) les espèces (Laurin & Cantino 2007), entre autres parce que de nombreux concepts d'espèces sont actuellement utilisés et que sous plusieurs d'entre eux, les espèces ne forment pas forcément des clades (de Queiroz 2007). Ces noms continueront donc à être régis par les codes Linnéens-Stricklandiens.

7. Conclusions et perspectives

Les opposants à la nomenclature phylogénétique prétendent que le passage vers un nouveau type de nomenclature générerait trop de confusion pour être profitable (Nixon *et al.* 2003). Cet argument est peu crédible car le système actuel permet un nombre très élevé de sens pour

chaque nom de coexister et permet à chaque clade de posséder de nombreux noms (Figures 1, 2). Peut-on imaginer un système plus confus ? Le PhyloCode permettra tout simplement de stabiliser le sens des noms de taxons en choisissant une des acceptions les plus fréquemment utilisées pour chaque nom, ce qui peut difficilement créer de la confusion.

Une comparaison avec l'informatique peut illustrer la relation entre confusion et progrès. Au début des années 1980, DOS (Disk Operating System) dominait largement le marché des micro-ordinateurs. Puis, Apple inventa le MacOS (en 1984), muni d'un interface graphique. IBM et Microsoft lancèrent peu après leurs propres systèmes d'exploitation munis d'interfaces graphiques (OS/2 et Windows). Cette révolution de l'interface graphique généra plus de confusion chez les millions d'utilisateurs d'ordinateurs du monde entier et coûta bien plus que les sommes mobilisables pour n'importe quelle révolution en nomenclature biologique. Pourtant, aujourd'hui, qui voudrait retourner à DOS ? Cette comparaison montre également que seul un système alternatif comportant des avantages importants et remportant un assez grand succès (comme Windows et le MacOS) peut générer de la confusion chez les utilisateurs des autres systèmes. Les autres systèmes (comme OS/2) n'ont pas assez d'impact pour créer une telle confusion.

Ainsi, seul un assez grand succès du PhyloCode lui permettrait de créer de la confusion chez certains adeptes de la nomenclature Linnéenne-Stricklandienne qui ne se seraient pas familiarisés avec les principes de base de la nomenclature phylogénétique. Or, pour qu'un tel nombre de systématiciens adopte un nouveau système, il faudrait qu'il comporte des avantages réels, car l'inertie favorise toujours le système déjà en place. L'avenir nous dira bientôt si le PhyloCode révolutionnera (ou non) la nomenclature biologique.

Finalement, que reste-t-il de Linné dans la nomenclature d'aujourd'hui ? Les rangs existent toujours, mais pour combien de temps ? Si la nomenclature phylogénétique est adoptée, ils joueront un rôle de moins en moins important. La nomenclature binominale semble devoir persister bien plus longtemps. Même si des propositions alternatives ont été faites (Cantino *et al.* 1999 ; Dayrat *et al.* 2004), elles n'ont pour l'instant pas obtenu le soutien d'une proportion suffisamment grande de la communauté scientifique pour avoir un impact. Ce système de noms binominaux est sans doute la contribution la plus importante et la plus durable de Linné à la nomenclature biologique.

Remerciements

Je remercie Christophe Roche et le comité d'organisation du congrès TOTh 2009 pour l'invitation et le financement qui me permirent de présenter la conférence inaugurale que cette contribution résume et David Marjanovic pour une relecture critique.

Bibliographie

- Béthoux O. (2007) : Propositions for a character-state-based biological taxonomy. *Zoologica Scripta*, vol. 36, N° 4, pp. 409–416
- Bottéro J. (1994) : *Babylone : à l'aube de notre culture*, Gallimard, Paris
- Bryant H. N. (1994) : Comments on the phylogenetic definition of taxon names and conventions regarding the naming of crown clades. *Systematic Biology*, vol. 43, N° 1, pp. 124–130.
- Calberg-Challot M., Lerat P., Roche C. (à paraître) : "Quelle place accorder aux corpus dans la construction d'une terminologie ?", Actes de la troisième conférence TOTh 2009, Terminologie & Ontologie : Théories et applications, Christophe Roche éd., Annecy, Institut Porphyre
- Cantino P. D., Bryant H. N., de Queiroz K., Donoghue M. J., Eriksson T., Hillis D. M. et Lee M. S. Y. (1999) : Species names in phylogenetic nomenclature. *Systematic Biology*, vol. 48, N° 4, pp. 790–807
- de Queiroz K. (2007) : Species concepts and species delimitation. *Systematic Biology*, vol. 56, N° 6, pp. 879–886
- de Queiroz, K. et Gauthier J. (1990) : Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Systematic Zoology*, vol. 39, N° 4, pp. 307–322
- Dayrat B., Schander C. et Angielczyk K. D. (2004) : Suggestions for a new species nomenclature. *Taxon*, vol. 53, N° 2, pp. 485–491
- Dubois A. (2006) : Incorporation of nomina of higher-ranked taxa into the International Code of Zoological Nomenclature: some basic questions. *Zootaxa*, vol. 1337, pp. 1–37
- Dubois A. (2007) : Naming taxa from cladograms: a cautionary tale. *Molecular Phylogenetics and Evolution*, vol. 42, N° 2, pp. 317–330
- Ereshefsky M. (2007) : Foundational issues concerning taxa and taxon names. *Systematic Biology*, 56, (2) pp. 259–301
- Frost D. R., Grant T., Faivovich J., Bain R. H., Haas A., Haddad C. F. B., de Sá R. O., Channing A., Wilkinson M., Donnellan S. C., Raxworthy C. J., Campbell J. A., Blotto B., Moler P., Drewes R. C., Nussbaum R. A., Lynch J. D., Green D. M. et Wheeler W. C. (2006) : The amphibian tree of life. *Bulletin of the American Museum of Natural History*, vol. 297, N° pp. 1–370

- Gauthier J. (1986) : Saurischian monophyly and the origin of birds, in Padian, K. (eds.), *The Origin of Birds and the Evolution of Flight*, California Academy of Sciences, San Francisco, pp. 1–55
- Gradstein F. M., Ogg J. G. et Smith A. G. (2004) : *A Geologic Time Scale 2004*, Cambridge University Press, Cambridge
- Hillis D. M. (2007) : Constraints in naming parts of the Tree of Life. *Molecular Phylogenetics and Evolution*, vol. 42, N° 2, pp. 331–338
- Laurin M. (1991) : The osteology of a Lower Permian eosuchian from Texas and a review of diapsid phylogeny. *Zoological Journal of the Linnean Society*, vol. 101, N° 1, pp. 59–95
- Laurin M. (2005) : The advantages of phylogenetic nomenclature over Linnean nomenclature, in Minelli, A. et Ortalli, G. et Sanga, G. (eds.), *Animal Names*, Istituto Veneto di Scienze, Lettere ed Arti, Venice, pp. 67–97
- Laurin M. (2008a) : Le PhyloCode, in Prat, D. et Roguenant, A. et Raynal-Roques, A. (eds.), *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*, Belin, Paris, pp. 411–420
- Laurin M. (2008b) : The splendid isolation of biological nomenclature. *Zoologica Scripta*, vol. 37, N° 2, pp. 223–233
- Laurin M. et Cantino P. D. (2004) : First international phylogenetic nomenclature meeting: a report. *Zoologica Scripta*, vol. 33, N° 5, pp. 475–479
- Laurin M. et Cantino P. D. (2007) : Second meeting of the International Society for Phylogenetic Nomenclature: a report. *Zoologica Scripta*, vol. 36, N° 1, pp. 109–117
- Laurin M. et Bryant H. N. (2009) : Third Meeting of the International Society for Phylogenetic Nomenclature: a Report. *Zoologica Scripta*, vol. 38, N° 3 pp. 333–337
- Lee M. S. Y. et Skinner A. (2007) : Stability, ranks, and the PhyloCode. *Acta Palaeontologica Polonica*, vol. 52, N° 3, pp. 643–650
- Minelli A. (2000) : The ranks and the names of species and higher taxa, or a dangerous inertia of the language of natural history, in Ghiselin, M. T. et Leviton, A. E. (eds.), *Cultures and Institutions of Natural History : Essays in the History and Philosophy of Sciences*, California Academy of Sciences, San Francisco, pp. 339–351
- Nixon K. C., Carpenter J. M. et Stevenson D. W. (2003) : The PhyloCode is fatally flawed, and the “Linnean” System can easily be fixed. *The Botanical Review*, vol. 69, N° 1, pp. 111–120
- Padian K. (1999) : Charles Darwin’s views of classification in theory and practice. *Systematic Biology*, vol. 48, N° 2, pp. 352–364
- Rieppel O. (2005) : Monophyly, paraphyly, and natural kinds. *Biology and Philosophy*, vol. 20, pp. 465–487
- Rowe T. et Gauthier J. (1992) : Ancestry, paleontology, and definition of the name Mammalia. *Systematic Biology*, vol. 41, N° 3, pp. 372–378

La nomenclature biologique aujourd'hui : que reste-t-il de Linné ?

Strickland H. E., Henslow J. S., Phillips J., Shuckard W. E., Richardson J. B., Waterhouse G. R., Owen R., Yarrell W., Jenyns L., Darwin C., Broderip W. J. et Westwood J. O. (1842) : Report of a committee appointed "to consider of the rules by which the Nomenclature of Zoology may be established on a uniform and permanent basis". *Annals and Magazine of Natural History*, 11, pp. 1-17

Strickland H. E., Henslow J. S., Phillips J., Shuckard W. E., Richardson J. B., Waterhouse G. R., Owen R., Yarrell W., Jenyns L., Darwin C., Broderip W. J. et Westwood J. O. (1843) : Series of propositions for rendering the nomenclature of zoology uniform and permanent, being the Report of a Committee for the consideration of the subject appointed by the British Association for the Advancement of Science. *Annals and Magazine of Natural History*, vol. 11, N° pp. 259-275

Voultsiadou E. et Vafidis D. (2007) : Marine invertebrate diversity in Aristotle's zoology. *Contributions to Zoology*, vol. 76, N° 2, pp. 103–120

Wolsan M. (1993) : Phylogeny and classification of early European Mustelida (Mammalia : Carnivora). *Acta Theriologica*, vol. 38, N° 4, pp. 345-384

A propos des auteurs

Michel Laurin

UMR 7207 – Centre de Recherches sur la Paléobiodiversité et les Paléoenvironnements

Muséum National d'Histoire Naturelle, 43 rue Buffon, CP 48

75005 Paris

michel.laurin@upmc.fr

http://tothweb.org/notes/?note_id=3669