



HAL
open science

Évaluer la santé de la taxonomie zoologique : histoire, méthodes et enjeux contemporains

Elise Tancoigne

► **To cite this version:**

Elise Tancoigne. Évaluer la santé de la taxonomie zoologique : histoire, méthodes et enjeux contemporains. Systématique, phylogénie et taxonomie. Museum national d'histoire naturelle - MNHN PARIS, 2011. Français. NNT : . tel-00707531

HAL Id: tel-00707531

<https://theses.hal.science/tel-00707531>

Submitted on 12 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Résumé

Ce travail est la première approche de la taxonomie zoologique basée sur les données de la plus grande base de données bibliographiques de zoologie, le *Zoological Record*, à l'aide d'outils scientométriques : méthodes cartographiques, comptages. Il remet en cause la conception selon laquelle la taxonomie serait menacée d'extinction, à l'instar de ses objets d'étude. Dans le contexte actuel de crise de la biodiversité, ce n'est pas une forme de déclin qui se révèle, mais plutôt une insuffisance des forces mises en jeu pour connaître l'ensemble de la biodiversité avant sa disparition. Ce travail sur les publications permet de donner une définition claire de cette discipline, et ainsi de proposer une nouvelle approche d'évaluation de sa santé. Puisque la taxonomie est définie comme *la science qui travaille sur des collections naturalistes à des fins nomenclaturales*, il est pertinent de proposer que sa santé soit évaluée en fonction du dynamisme de ses collections. Le déclin de la taxonomie n'étant plus considéré comme une donnée factuelle, il mériterait d'être abordé en tant que discours tenu par une communauté scientifique face à de profondes modifications de ses pratiques et face à la disparition de ses objets d'étude. Une approche relevant des sciences humaines traiterait sans doute avec profit cette thématique.

Mots-clés : taxonomie, scientométrie, *Zoological Record*, crise de la biodiversité, déclin, collections

Abstract

This work is the first approach to zoological taxonomy based on data from the largest bibliographic database of Zoology, the Zoological Record, analyzed through scientometric tools : science mapping and countings. It challenges the view that taxonomy would be an endangered field, as its objects of study are. In the current biodiversity crisis, this is not a form of decline, but rather a lack of involvement to determine the overall biodiversity before its disappearance. This work on publications provides a clear definition of the discipline, and so proposes a new approach to assess its vitality. Since taxonomy is defined as the branch of science that works on collections with nomenclatural aims, it is pertinent that its health should be evaluated according to the dynamism of its collections. The decline of taxonomy no longer being regarded as a factual matter, it deserves to be addressed as a discourse given by a scientific community facing significant changes in both its practices and the disappearance of its objects of study. The evaluation of the vitality of taxonomy, as a theme, would no doubt benefit from a Humanities-orientated approach.

Keywords : taxonomy, scientometry, *Zoological Record*, biodiversity crisis, decline, collections

Remerciements

Je tiens à remercier le Muséum national d'Histoire naturelle, les membres de la commission d'attribution des allocations de l'année 2008, et Alain Dubois pour m'avoir accordé un lopin de terre où faire germer une *Thesis elisae*;

Alain Dubois à nouveau, pour m'avoir permis de partir explorer les cabanes des jardiniers alentour;

Anne Sigogneau, pour m'avoir entrouvert les portes de ces cabanes et accompagnée avec sollicitude depuis notre rencontre en master, sans se décourager de jouer les piverts;

Alessandro Minelli, Jean-Yves Rasplus, Jean-Philippe Cointet et Pascal Cassy pour avoir accepté de se pencher sur ce nouveau spécimen;

Nigel Robinson pour m'avoir aidée à déterminer la nature du sol;

Mathilde de Saint Léger, Sébastien Heymann pour m'avoir transmis le maniement de la serfouette et de la serpette;

Régine Vignes-Lebbe, Donato Bergandi, Pascal Cassy, Philippe Grandcolas pour m'avoir mis en garde contre les OGM;

Pascal Cristofolli, Viera Rebelledo-Dhuin, Jean-Philippe Cointet, Marc Barbier, Philippe Breucker, Murat Sariyar et Pees Jan Van Eck pour nos échanges horticoles;

le laboratoire des Reptiles et Amphibiens - Annemarie, Myriamne, Victoire, Laure, Jérôme, Yves, Marc, Emmanuel, Didier, Stéphane, Sandra, Ivan - pour l'oxygène que vous m'avez apporté - devrais-je dire le CO₂?;

la cellule psy du département HNS : Florence, Aline, Maya, Alix, Erik, Nicolas, Cindy, Anneclaire, Jonathan, Aurélie, Marine, Jean-Cristan, Rihlat, Noémie, Romain, Emeric, Emmanuel... pour leur aide à l'élagage et leurs réserves d'eau;

pour les nutriments et éléments rares, je remercie Cyprien Bole, statisticien diphonique; Florent Gibaux, moissonneur; Julio Mas, programmeur passionné et Aurélien Miralles, bibliothécaire insoupçonné;

sans oublier : la BDEMteam 2009-2010 : Noëlie, Lorène, Aude, Jeanne, Chloé, Erik, Jocelyn, Bouziane, Yann - vous reprendrez bien un peu de jus de tomate? ainsi que toutes les personnes avec qui j'ai pu collaborer pendant mon monitorat, et en particulier Julie Marin, Catherine Morge, Adèle Corvez, Pascale Gendrot, François Jacquet et Anaïs Grand.

Pour la dernière saison je remercie tous les amis qui m'ont écoutée parler des intempéries et ravageurs rencontrés sur mon chemin et/ou qui ont évité de me demander si mon sujet avait quelque chose à voir avec la scientologie... un merci plus particulier à Carz et António pour leurs formules percutantes au plus froid de l'hiver;

Romain Thomas, Régis et Julie Lebloup, qui ont tracé ce chemin de traverse conduisant les promeneurs au jardin;

Loizic, Didier et Claude pour avoir osé l'emprunter les premiers;

ainsi que, pour finir, toutes les personnes qui m'ont encouragée à poursuivre dans la recherche depuis mon premier stage à la fac du Mans lorsque j'étais en 6^e : Nirina Randrianantoandro, Alain Rojo de la Parz, Nathalie Guivarc'h, Nadine Imbault, Vincent Burlat, Odile Poncy, Dario de Franceschi, Anaïs Boura

... et les trois lecteurs préférés de Grégoire Solotareff...

Sommaire

Préambule	xi
I « Mesurer une discipline » : l’affaire de la scientométrie	1
1 Une brève histoire de la scientométrie	3
1.1 Les précurseurs	4
1.1.1 Candolle, Galton et Cattell	4
1.1.2 Naissance de la bibliométrie	5
1.2 Les fondateurs	5
1.2.1 Solla Price : mise en place d’une théorie sociologique	5
1.2.2 Garfield : mise en place des outils	6
1.3 L’essor d’une discipline	6
1.3.1 Développement des outils	6
1.3.2 Reconnaissance et visibilité de la discipline	6
1.3.3 Le volet applicatif : l’évaluation de la recherche et la veille technologique	7
2 Techniques de la scientométrie	9
2.1 Distributions bibliométriques : Lotka, Bradford, Zipf	9
2.2 Indicateurs univariés	11
2.3 Cartes relationnelles de première génération	13
2.3.1 Co-publication	13
2.3.2 Citation de revues à revues, réseaux de citation	13
2.3.3 Analyse des co-citations	14
2.4 Cartes relationnelles de seconde génération	14
3 Le <i>Zoological Record</i>, base d’étude privilégiée	17
3.1 Présentation	18
3.1.1 Historique	18
3.1.2 Propriétés	19
3.1.3 Informations techniques	22
3.1.4 Distribution des articles dans le <i>ZR</i>	26
3.1.5 Outils d’analyse du <i>Zoological Record</i>	28

3.2	Construction d'un corpus de travail issu du <i>ZR</i>	29
3.2.1	Faut-il chercher par genres ou par familles?	29
3.2.2	Faut-il faire une recherche avec des termes complets ou utiliser les troncatures?	30
3.2.3	Faut-il rechercher par les noms vernaculaires?	33
3.2.4	Requête finale	33
3.3	Gestion des auteurs synonymes	33
3.3.1	Quelle méthode utiliser?	35
3.3.2	Outils à disposition	36
3.3.3	Algorithme d'homogénéisation	36
II	La taxonomie, une discipline isolée	43
4	Une discipline aux contours flous	45
4.1	Étymologie	45
4.1.1	Création	45
4.1.2	Taxonomie ou taxinomie?	46
4.2	Définitions d'auteurs	46
4.3	Définition par la pratique	50
4.4	Absence de consensus	51
4.5	Vers une cartographie de la zoologie	52
5	Cartographies de la zoologie	59
5.1	Méthodologie	59
5.1.1	Principe général	59
5.1.2	Outils disponibles	61
5.1.3	Réalisation des cartes	62
5.2	Résultats	63
6	Cartographie de citations au sein de la zoologie	73
6.1	Méthodologie	73
6.1.1	Choix du corpus et du <i>WoS</i>	73
6.1.2	Estimation de la couverture taxonomique du <i>WoS</i>	74
6.1.3	Définition des mots-clés de recherche	76
6.1.4	Construction du graphe	77
6.2	Résultats	78
III	Un déclin relatif	83
7	Que mesurer?	85
7.1	La notion de déclin	85

SOMMAIRE

7.2	Définition du corpus	86
8	Mesures des productions et acteurs	89
8.1	Nombre de publications	89
8.1.1	Compte de présence	89
8.1.2	Compte fractionné	91
8.2	Nombre d'espèces et sous-espèces décrites	93
8.2.1	Méthodologie	93
8.2.2	Résultats	93
8.3	Nombre de taxonomistes	93
8.3.1	Méthodologie	93
8.3.2	Résultats	94
8.4	Nombre de revues	94
8.4.1	Méthodologie	94
8.4.2	Résultats	95
9	Discussion	103
9.1	Absence de déclin global	103
9.2	Vers une redistribution des forces ?	104
IV	Le déclin, une inquiétude récurrente	107
10	Une inquiétude récurrente dans <i>Nature</i> et <i>Science</i> depuis 1946	109
10.1	Méthodologie	109
10.2	Résultats	110
10.2.1	L'après-guerre	110
10.2.2	Émergence de nouvelles techniques en taxonomie	111
10.2.3	Remise en cause des objectifs	111
10.2.4	Remise en cause des pratiques	112
10.2.5	L'ère informatique	113
10.2.6	L'ère du barcode	114
10.2.7	Conclusion	114
11	Des mesures depuis le début du XX^e siècle	117
11.1	Une préoccupation dès le début du XX ^e siècle	118
11.2	Des arguments récurrents	119
V	Faut-il développer la taxonomie ?	123
12	Importance de la taxonomie	125
12.1	Pourquoi développer la taxonomie ?	125
12.2	Mesure de l'apport de la taxonomie aux autres disciplines	126

12.2.1	Méthodologie	127
12.2.2	Résultats	128
12.2.3	Conclusion	130
13	Proposition de mesures de sa santé	131
13.1	Quand <i>inertie</i> remplace <i>déclin</i>	131
13.2	Évaluer la taxonomie à travers ses collections	134
	Conclusion	137
	Annexes	141
A	Article de <i>Frontiers in Zoology</i>	141
B	Détail des champs du <i>ZR</i>	155
C	Exemples de scripts de traitement des données sous R	161
C.1	Fonction d'import des notices	161
C.2	Fonction de déduplication des notices	162
D	Définitions de la taxonomie dans la littérature	165
D.1	Quelques définitions des termes <i>taxinomie</i> , <i>taxonomie</i> ou <i>taxonomy</i> .	165
D.2	Quelques définitions des termes <i>sysématique</i> ou <i>systematics</i>	169
E	Pas à pas de l'historiographe	171
F	Description et période de validité des 43 termes de premier niveau du <i>subject thesaurus</i> du <i>ZR</i>	173
G	Thesaurus correspondant au terme <i>Systematics</i> du <i>ZR</i>	179
H	Stratégie de recherche bibliographique sur le thème de la santé de la taxonomie	185
	Bibliographie	187
	Liste des figures	207
	Liste des tableaux	211

Préambule

Le déclin de la taxonomie, un thème consensuel

Taxonomists are an endangered group
Pearson *et al.* (2011, p. 58)

*Scientists who classify, describe and examine
the relationships between organisms
are themselves going extinct*
McClain (2011)

Récemment, deux articles parus dans des médias de nature différente, le journal *BioScience* (Pearson *et al.*, 2011) et la série périodique *From the Fields* issue du réseau de blogs scientifiques *Wired Science* (McClain, 2011) ont souhaité attirer l'attention de la communauté scientifique sur le déclin de la taxonomie, et la disparition des taxonomistes : *Recovery Plan for the Endangered Taxonomy Profession* et *The Mass Extinction of Scientists Who Study Species*.

Plus récemment encore, une centaine d'E-mails ont été échangés sur la liste de diffusion Taxacom à la suite de la question *Why Taxonomy does NOT matter ?* posée fin avril 2011 sur cette même liste par le taxonomiste allemand Fabian Haas¹.

Depuis la reconnaissance officielle, en 1992, d'un « handicap taxinomique », c'est-à-dire d'un « *manque d'infrastructures, de moyens et de financements* » (Anonymous, 1992) dans cette discipline, un certain nombre de projets internationaux ont vu le jour : GBIF (*Global Biodiversity Information Facility*), EDIT (*European Distributed Institute of Taxonomy*), PEET (*Partnerships for Enhancing Expertise in Taxonomy*), BioNET... Ces différents projets d'aide à la taxonomie continuent de recevoir des crédits et un soutien administratif sous une forme ou sous une autre².

1. <http://mailman.nhm.ku.edu/pipermail/taxacom/2011-April/072162.html>.

2. <http://vbrant.eu/>

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5451

<http://www.gbif.org/governance/how-to-join/>

<http://www.bionet-intl.org/opencms/opencms/index1.jsp>.

Il semble donc : d'une part, qu'il existe aujourd'hui un consensus autour de l'idée que la taxonomie est une discipline en danger ; et, d'autre part, que cette idée soit aussi défendue par des non-taxonomistes. Les deux articles aux intitulés inquiétants précédemment mentionnés ont en effet pour auteurs un écologiste et biologiste de la conservation (David L. Pearson), un historien des sciences (Andrew L. Hamilton), un coléoptériste (Terry L. Erwin) et le directeur scientifique adjoint du National Evolutionary Synthesis Center (Craig McClain).

Pour preuve que ce propos semble bien ancré, aucun des acteurs précédemment cités ne fournit de données argumentant ce déclin de la taxonomie. Si le premier article aborde le problème sous un angle historique, en s'appuyant sur les travaux d'historiens, de taxonomistes ou de philosophes, le second article construit son argumentation à partir d'une étude de cas (l'état des lieux des ressources taxonomiques disponibles pour le groupe des Aplacophores), de témoignages de collègues et de la perception qu'a son auteur des retombées de grands programmes de soutien à la taxonomie (*My personal experience highlights how [...]*). Quant aux grands projets internationaux, ils présentent ce « *manque d'infrastructures, de moyens et de financements* » comme une donnée de base évidente, à l'origine de leurs actions :

Objectives : "To set up the means, tools and infrastructure to produce a more rational and a more effective framework for European biodiversity research" (Vibrant)

the National Science Foundation seeks to enhance taxonomic research and help prepare future generations of experts (PEET)

Because human threats to biodiversity occur across large spatial and temporal scales, biodiversity and ecosystem monitoring, forecasting, and risk assessments require data to be organised in a globally-accessible, integrated infrastructure (GBIF)

Our work contributes to raising awareness of the importance of taxonomy to society, building and sharing of capacity, and meeting taxonomic needs via innovative tools and approaches (BioNET)

Or, si la reconnaissance officielle du manque de moyens de la taxonomie est ancienne (Anonymous, 1992) et les articles proposant des solutions nombreux (par ex. Bisby *et al.*, 2002; Gewin, 2002; Godfray, 2002b; Knapp *et al.*, 2002; Thiele et Yeates, 2002; Barratt *et al.*, 2003; Dunn, 2003; Rodman et Cody, 2003; Scoble, 2004; Polaszek *et al.*, 2005; Crisci, 2006a; Evenhuis, 2007; Penev *et al.*, 2009; Wheeler, 2009), il est extrêmement difficile de trouver des études faisant un état des lieux global de la discipline.

PRÉAMBULE

Il existe, en effet, quelques états des lieux nationaux ; récents (Tillier *et al.*, 2000; Herbert *et al.*, 2001; Klopper *et al.*, 2002; Shimura, 2003; Mather, 2004; Australian Government department of the Environment Water Heritage and the Arts, 2006; Marques et Lamas, 2006; Michán *et al.*, 2008; Smith *et al.*, 2008; House of Lords Science and Technology Committee, 2008; Michán et Llorente Bousquets, 2009) ou plus anciens (Advisory Board for the Research Councils, 1979; Valdecasas *et al.*, 1992; Feinsinger, 1996). Toutefois, il n'existe pas d'initiative à grande échelle. Lorsque ces états des lieux ne se limitent pas à un pays, c'est qu'ils sont réalisés à l'échelle d'un taxon (Coomans, 2002; Loeb et Leschen, 2005; Bouchet et Duarte, 2006; Rodrigues *et al.*, 2010, par ex.).

Demande de quantification globale pour la discipline

Ce travail a donc pris racine dans une volonté de disposer d'outils de mesure pour pouvoir décrire, à grande échelle et avec des outils quantitatifs, le paysage et les avancées de la taxonomie zoologique. Il visait ainsi à :

1. comprendre l'évolution des pratiques taxinomiques au cours du temps, en fonction des groupes étudiés, des journaux scientifiques et des pays ;
2. mieux cerner, de manière quantifiée, la réalité de la pratique taxinomique ;
3. formuler des recommandations pour son adéquation aux besoins scientifiques et sociétaux ;
4. mettre en évidence des critères d'analyse et d'évaluation pour un meilleur suivi des progrès de la discipline et donc de la pertinence, de l'efficacité et des retombées dans ce domaine des grands programmes internationaux.

État de l'art

Les travaux quantitatifs sur l'état de la taxonomie font appel à différents procédés. Leurs données peuvent être issues de réponses à un questionnaire (Klopper *et al.*, 2002; Shimura, 2003; Australian Government department of the Environment Water Heritage and the Arts, 2006; House of Lords Science and Technology Committee, 2008), de bases de données bibliographiques spécialisées (Marques et Lamas, 2006; Michán *et al.*, 2008; Michán et Llorente Bousquets, 2009), ou, plus rarement, du dépouillement d'un échantillonnage d'articles de taxonomie (Bouchet et Duarte, 2006) ou encore de l'évolution de la participation à des associations naturalistes (Smith *et al.*, 2008).

Dans le cadre de notre problématique, l'étude des données de sociétés naturalistes ne nous aurait pas permis d'étudier la taxonomie à une échelle mondiale, celles-ci étant toujours restreintes à un taxon ou une région géographique donnée. L'étude par questionnaire, quant à elle, ne nous permettait pas d'aborder la dimension temporelle souhaitée. Deux approches pouvaient donc être retenues : l'approche par dépouillement d'un échantillon d'articles, et l'approche basée sur les bases de données spécialisées.

L'approche par dépouillement a tout d'abord été envisagée, à la suite d'un précédent travail de Master 2, *Métataxonomie : l'état de la taxonomie zoologique à l'aube du XXI^e siècle* (Tancoigne, 2008). Cette approche consistait à :

1. échantillonner aléatoirement un certain nombre d'articles de taxonomie zoologique ;
2. extraire de ces articles les données sur :
 - les principales caractéristiques des travaux de recherche en taxonomie : techniques employées pour récolter, étudier, caractériser et décrire les nouveaux taxons ;
 - les taxons étudiés : groupe taxinomique ; taxons nouveaux ou non ; rang nomenclatural ; catégorie taxinomique ; etc. ;
 - les spécimens étudiés : date de récolte ; provenance (région du globe, pays, latitude, longitude, altitude, profondeur) ; nombre ; sexe, stade et taille ; lieu de conservation (région du globe, pays, ville, institution ou privé, inconnu) ; etc. ;
 - les auteurs : nombre ; nationalité ; sexe ; âge ; taxinomiste professionnel, étudiant, retraité ou amateur ; type d'établissement de travail ; pays et ville d'origine et de travail ; origine sociologique (profession des parents, importance de la fratrie) ; durée totale de la période d'activité en taxonomie et nombre de publications produites ; groupes taxinomiques étudiés ; régions du globe et pays étudiés ; etc. ;
 - les publications scientifiques en taxonomie : type de publication : monographie, révision, descriptions isolées de nouveaux taxons, modifications taxinomiques et nomenclaturales ponctuelles des travaux antérieurs (mises en synonymie, résolution d'homonymies, etc.) ; etc. ;
 - les descriptions : structure ; informations fournies ; tableaux ; iconographie ; etc. ;
 - les noms donnés aux taxons : signification ; étymologie ; structure ; nombre de lettres et syllabes ; statut grammatical ; relation avec la langue de l'auteur ; « qualité » en matière de communication internationale optimale à long terme (euphonie, signification, dimension mnémotechnique, longueur, etc.) ;
3. ordonner ces données dans une base de données relationnelles dédiée ;

PRÉAMBULE

4. explorer ces données par des moyens statistiques pour en dégager une description et des tendances.

L'approche par échantillonnage est une pratique qui fait sens en biologie, mais également en histoire (Lemercier et Zalc, 2008). Échantillonner des données dans une perspective historique est donc réalisable, à condition que l'échantillon soit représentatif; c'est-à-dire qu'il représente la population dont il est issu. Dans notre cas, nous souhaitons réunir des informations sur les pratiques de recherche en taxonomie, les taxons étudiés, les spécimens étudiés, les taxonomistes, les publications et les noms donnés aux taxons; et mettre en opposition différents sous-groupes d'individus au sein de chacune de ces catégories. Par conséquent, il fallait avoir un échantillon qui soit représentatif des différents sous-groupes de chacune de ces catégories. Or, l'approfondissement du travail de M2, qui donna lieu à une publication (voir annexe A), nous apprit qu'il n'était pas possible de généraliser à l'ensemble de la taxonomie les résultats que nous avons obtenus sur un échantillon d'articles de 2006-2007 issus de *Zootaxa*, sauf si nous regroupions nos données sous des catégories très larges : Vertebrata, Arthropoda et Other pour les taxons; regroupement des continents en zones géographiques plus étendues... Pourtant, ce travail était le fruit d'un dépouillement de 748 articles, réalisé en plusieurs mois alors même que chacun des articles était directement disponible puisque accessible en ligne sous format PDF.

Il était donc clair que, dans ces conditions, il n'était pas envisageable d'utiliser la même approche pour avoir un aperçu précis de 250 ans d'activité taxonomique. Les difficultés d'accès au contenu intégral d'articles échantillonnés aléatoirement, le temps nécessaire à leur analyse étaient des investissements disproportionnés au regard de la précision qui aurait pu être obtenue dans les résultats.

En outre, si analyser finement des publications s'avérait pertinent pour connaître les pratiques de recherche en taxonomie, les taxons étudiés et les publications, cela ne se révélait pas la meilleure porte d'entrée pour répondre aux questions que nous nous posons sur les spécimens étudiés, les taxonomistes, et les noms donnés aux taxons. Pour toutes ces questions, en effet, il semblait plus pertinent d'avoir accès directement aux informations via les bases de données dédiées, à savoir :

bases de données de collection : l'informatisation d'un nombre croissant de collections d'histoire naturelle donne aujourd'hui accès à un grand nombre de données de spécimens (lieu de collecte, lieu de conservation, etc.). Des organisations internationales comme le GBIF travaillent sur la centralisation de toutes ces données sur la biodiversité (Edwards *et al.*, 2000);

annuaires de systématiciens : il existe aujourd'hui un annuaire des taxonomistes qui recense plus de 4 900 d'entre eux : *the World Taxonomist Database* (alimenté par inscription volontaire);

bases de données nomenclaturales : nous pouvons là encore citer des bases importantes, recensant les noms de la littérature zoologique : [uBio](#), [Index to Organism Names](#), [ITIS](#), [SysTax](#) ou encore [Taxonomicon](#).

Pour finir, ce sujet contenait en filigrane un certain nombre de présupposés :

- qu’il existe une discipline clairement définie appelée *taxinomie* ;
- que cette discipline est nécessaire pour la science et la société ;
- qu’elle doit être développée prioritairement aujourd’hui ;
- que l’histoire de la taxinomie peut être appréhendée de manière pertinente par une approche quantitative sur le contenu de ses publications ;
- que la connaissance de cette histoire permettra de mettre en place des indicateurs d’évaluation de sa santé.

Pour toutes ces raisons, j’ai choisi de me focaliser dans ce travail sur l’étude de ces présupposés, en utilisant pour cela la plus grande base de données bibliographique de zoologie, le *Zoological Record*, dont la couverture s’étend de 1864 à aujourd’hui.

Organisation du document

Nous allons voir en premier lieu que la question épineuse de la mesure d’une discipline scientifique n’est pas nouvelle. Elle est l’objet d’une discipline à part entière, la scientométrie, avec sa propre histoire et ses propres développements. Je présenterai brièvement ses principales méthodes : l’approche par comptages et l’approche relationnelle. Je détaillerai ensuite la principale base de données utilisée dans la suite de ce travail, le *Zoological Record*, ainsi que les outils que j’ai été amenée à mettre en place pour la manipulation de ses données.

Dans la seconde partie de ce travail nous verrons qu’il n’existe pas de définition consensuelle de ce qu’est la taxinomie. Je mobiliserai l’approche scientométrique relationnelle pour délimiter la taxinomie et en proposer ainsi une nouvelle définition. Nous verrons que c’est une discipline thématiquement isolée des autres disciplines zoologiques.

La troisième partie de ce travail appliquera l’approche scientométrique par comptage au *Zoological Record* pour déterminer l’ampleur du déclin de la discipline au cours du temps. Je m’intéresserai plus particulièrement au nombre de travaux, revues, espèces et sous-espèces décrits ; ainsi qu’au nombre d’auteurs en systématique. Nous verrons que le déclin tel qu’il est présenté dans la littérature n’est pas visible à travers ces chiffres excepté pour la période 1985-1995, et que le problème devrait

PRÉAMBULE

être considéré comme une résultante de la croissance des autres disciplines et d'une redistribution des forces taxonomiques.

En dernières parties je m'intéresserai à la façon dont est traitée la thématique du déclin dans la littérature. Nous verrons que cette thématique est ancienne, tout comme les tentatives de quantification de ses avancées, et qu'elle s'accompagne d'un argumentaire similaire quelle que soit l'époque. Je donnerai ensuite des justifications en faveur d'un développement plus important de la taxonomie.

En conclusion je proposerai qu'à l'avenir la taxonomie soit définie non pas à travers ses productions, ou ses buts, mais à travers son objet d'étude, la collection naturaliste. Dans une perspective d'évaluation de la vitalité de cette discipline, j'encouragerai la mise en place d'indicateurs basés sur l'utilisation et les forces de ces collections plutôt que sur les résultats qu'elle produit.

Première partie

« Mesurer une discipline » :
l'affaire de la scientométrie

Chapitre 1

Une brève histoire de la scientométrie

La scientométrie est une discipline qui s'est développée il y a une cinquantaine d'années et qui se définit comme *les recherches quantitatives de toutes les choses concernant la science et auxquelles on peut attacher des nombres* (Solla Price cité par Polanco (1995)). Elle est donc définie à travers son objet d'étude, la science, et sa méthodologie, une approche quantitative. Parce que cette définition ne tient pas compte des objectifs que la discipline se fixe, on lui en trouve plusieurs. Présents dès le départ, aux fondements et intérêts radicalement différents, ils distinguent d'une part, la scientométrie qui propose des outils de gestion de la recherche et d'autre part, la scientométrie qui cherche à comprendre le déroulement des sciences. Le développement de ces deux scientométries est étroitement lié, et il est impossible de parler de l'une sans évoquer l'autre. La première a été à l'origine du développement méthodologique qui a permis l'essor de la discipline ; c'est elle qui prédomine aujourd'hui en France. Leur importance relative a toutefois varié avec le temps et les pays.

Les deux approches ont en commun, outre la méthodologie, de s'intéresser à une analyse de la science. La scientométrie fut donc très rapidement questionnée d'un point de vue philosophique et sociologique, et ses questions convergèrent vers celles de disciplines existantes :

- la philosophie des sciences, ou épistémologie de la connaissance scientifique, discipline institutionnalisée à la fin du XIX^e siècle en Allemagne, qui construit des modèles « *pour élucider ce qui est essentiel dans les concepts, théories, méthodes et rapports mutuels des sciences établies* » (Ulises Moulines, 2006, p. IX) ;
- l'histoire des sciences, ou étude de l'évolution de la connaissance scientifique ;

- la sociologie des sciences, fondée par Robert K. Merton vers 1940, vouée au départ à l'étude des mécanismes de fonctionnement de la communauté scientifique en tant qu'institution humaine sans rentrer dans les contenus (Courtial, 2006);
- les *science studies*, champ de recherche interdisciplinaire qui vise à élucider le fonctionnement concret de la science et de son articulation avec le reste de la société, en s'appuyant tant sur la sociologie que sur la philosophie, l'économie, l'anthropologie, l'histoire ou d'autres disciplines des sciences humaines et sociales.

1.1 Les précurseurs

1.1.1 Candolle, Galton et Cattell

Essaie-t-on soi-même de pénétrer dans le labyrinthe des faits et des théories scientifiques, on est arrêté très vite par deux obstacles. L'un est l'immensité du sujet, l'autre le sentiment qu'il ne faut pas substituer sa propre opinion à celle du public scientifique, ni surtout à l'opinion des hommes spéciaux de chaque science, qui détermine presque toujours l'opinion générale des savants. Dans le but d'éviter ces difficultés, je me propose d'essayer une méthode qui n'a pas encore été employée. Au lieu de consulter les auteurs et de mêler à leurs jugements mes propres opinions, dont la valeur est nulle, excepté dans une seule science, je chercherai ce que les principaux corps savants de l'Europe ont pensé des hommes qui se sont distingués depuis deux siècles.

Alphonse de Candolle, p. 22 (1806-1893)

Coïncidence curieuse, c'est le fils du créateur du terme *taxonomie* Augustin Pyramus de Candolle (1778-1841) qui réalise avec son *Histoire des sciences et des savants depuis deux siècles*, ce que l'on considère être le premier travail de scientométrie. Son analyse du nombre de membres étrangers nommés par les principales sociétés ou académies des sciences européennes lui permet d'évaluer la contribution de chaque nation au développement scientifique mondial (Polanco, 1995).

Elle fut suivie par les travaux statistiques sur les scientifiques de Francis Galton (1822-1911), qui cherchait à montrer que l'hérédité est primordiale dans le choix d'une carrière et d'une réussite scientifique, et James McKeen Cattell (1860-1944), qui cherchait à démontrer que les conditions de travail des scientifiques en Amérique étaient moins bonnes que celles d'Europe et que le nombre de scientifiques dans le monde n'était pas assez élevé (Galton, 1874; Cattell, 1906 cités par Godin, 2005, p. 8).

1.2 Les fondateurs

Si ces travaux représentent les premières études statistiques portant sur la science, et en particulier les scientifiques, l'histoire de la scientométrie retient toutefois qu'elle s'est développée à partir de la bibliométrie, qui date elle-même du début du XX^e siècle.

1.1.2 Naissance de la bibliométrie

L'une des maladies de notre époque est la multiplicité des livres ; ils surchargent tellement le public que celui-ci est incapable de digérer l'abondance de matière oiseuse quotidiennement éclosée et répandue dans le monde
Barnaby Riche, 1613 ; cité par de Solla Price (1972, p. 68)

L'idée selon laquelle les scientifiques doivent faire face à une somme toujours plus vertigineuse de documents n'est pas nouvelle. Ce discours revient même de façon assez cyclique (Braun et Zsindely, 1985). Puisqu'il faut faire des choix, Cole et Eales (1917) proposent dès le début du XX^e siècle une méthode permettant d'identifier les journaux indispensables à lire dans leur domaine, basée sur le dénombrement de textes scientifiques. Puis c'est Gross et Gross (1927) qui utilisent à leur tour ces méthodes pour aider les bibliothèques à faire des choix dans leurs abonnements, celles-ci ayant un budget trop limité pour souscrire à l'ensemble des journaux existants. Avec eux naît donc l'application aux journaux scientifiques de la bibliométrie - étude statistique des productions écrites (Rostaing, 1996, p. 9) - déjà avancée sur des publications juridiques (Shapiro, 1992).

Les lois fondamentales de la bibliométrie sont énoncées rapidement, par Lotka (1926), Bradford (1934) et Zipf (1935). Le mot **bibliométrie** apparaîtra lui-même tardivement : « *application des mathématiques et des méthodes statistiques aux livres, articles et autres moyens de communication* » (Pritchard, 1969).

1.2 Les fondateurs

1.2.1 Solla Price : mise en place d'une théorie sociologique

Dans son ouvrage de référence *Little science, big science*, de Solla Price (1963) considéra que l'activité scientifique est régie par des règles sociologiques et proposa de nouvelles lois ainsi qu'un but différent à l'usage de la statistique sur les textes scientifiques. Avec lui naquit donc une *science de la science* (de Solla Price, 1972), dont les techniques furent appelées **scientométrie** pour les distinguer conceptuellement de celles de la bibliométrie, même si elles n'en différaient pas en pratique (Rostaing, 1996, p. 10).

1.2.2 Garfield : mise en place des outils

Les premiers instruments de travail de la scientométrie apparurent peu de temps après. Jusqu'alors, il n'existait pas de grande banque de données portant sur les productions scientifiques de domaines différents. Dès 1955, Eugene Garfield imagina pour l'aide à la recherche bibliographique des scientifiques l'existence d'un répertoire interdisciplinaire qui regrouperait les articles des principaux périodiques et leurs références (Garfield, 1955). Il fonda l'Institute for Scientific Information (ISI) en 1958 (Yancey, 2005) et la première version du Science Citation Index (SCI) vit le jour sous format papier à l'ISI en 1963. Les premiers instruments de travail de la scientométrie furent donc les journaux sélectionnés dans le SCI.

À l'Est comme à l'Ouest, la scientométrie s'inscrit lors de sa naissance dans une vision très positiviste de l'activité de recherche. Non seulement la science constitue une activité exemplaire, qui se distingue radicalement des autres activités humaines par son caractère hautement rationnel, mais de plus les méthodes qui fondent sa rigueur peuvent et doivent lui être appliquées : c'est en quelque sorte du scientisme à la puissance deux. (Callon et al., 1993, p. 6)

1.3 L'essor d'une discipline

1.3.1 Développement des outils

La scientométrie vit ensuite une grande avancée avec le développement des ordinateurs. Leur amélioration fit voir le jour aux premiers Systèmes de Gestion de Bases de Données (SGBD) à la fin des années 1960, mettant ainsi fin au stockage de l'information sous forme de fichiers simples. De grandes quantités de données pouvaient alors être stockées de manière organisée. Cela entraîna l'informatisation des bases de données bibliographiques et la création de nouvelles bases. Les chercheurs purent alors rapatrier ces informations dans leur laboratoire pour les étudier et de nouvelles méthodes, qualitatives, furent développées en Europe occidentale : analyse des mots associés, analyse des co-auteurs, analyse des co-opérations... (Rostaing, 1996).

1.3.2 Reconnaissance et visibilité de la discipline

En 1979, le lancement d'une revue intitulée *Scientometrics* réunit la *science de la science* de l'Ouest et la *naukovodemie* de l'Est (Callon et al., 1993, p. 3). En France, l'Association pour la mesure des sciences et des techniques (ADEST) fut

1.3 L'essor d'une discipline

créée en 1987 par Michel Callon. Au niveau international, l'International Society for Scientometrics and Informetrics (ISSI) fut créée en 1993.

1.3.3 Le volet applicatif : l'évaluation de la recherche et la veille technologique

Initialement imaginés [les indicateurs] comme une aide à la compréhension de l'évolution de la communauté scientifique et de ses paradigmes, les instances dirigeantes ont vu là l'occasion de disposer d'un système d'évaluation de la recherche. Rostaing (1996, p. 11)

Très rapidement, les grandes bases de données développées par l'ISI ont été vues comme un moyen de produire des indicateurs de la « vitalité » d'un chercheur, d'une institution, d'un pays ou d'un domaine de recherche, à des fins décisionnelles. L'un des débats qui agite la communauté scientifique actuellement, toutes disciplines confondues, est celui qui concerne l'évaluation de la recherche. Nous n'entrerons pas dans ce débat ici, mais il faut savoir que la question de la légitimité des indicateurs produits par l'ISI (propriété de Thomson Reuters depuis 1992) est toujours traitée vigoureusement, et que des alternatives existent (voir par exemple Bergstrom *et al.*, 2008; González-Pereira *et al.*, 2010). Un récent rapport de l'Académie des sciences (Académie des sciences, 2011) soulève ce problème.

Chapitre 2

Techniques de la scientométrie

2.1 Distributions bibliométriques : Lotka, Bradford, Zipf

Les travaux les plus anciens font appel à la notion de regroupement et dénombrement de travaux scientifiques sur des critères précis. On cherche à savoir, par exemple, la quantité de travaux publiés sur un sujet donné, le nombre de journaux d'un pays particulier, le nombre d'auteurs d'un domaine précis... Ces regroupements peuvent être classés entre eux, et donnent alors lieu à des distributions statistiques (Figure 2.1).

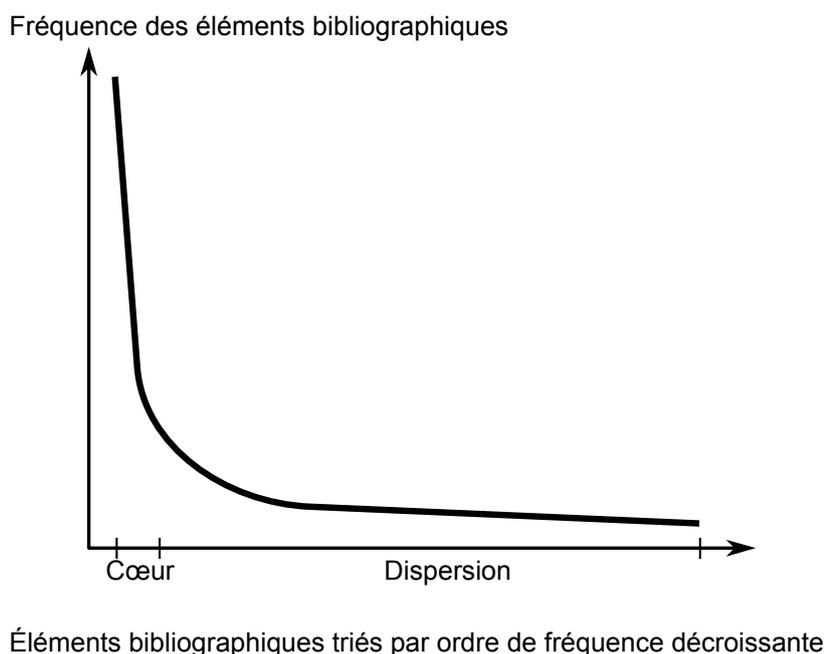
On observe très souvent la même distribution. Elle est caractérisée par une partie « cœur » et une partie « dispersion ». Par exemple, peu d'auteurs publient beaucoup, et beaucoup d'auteurs publient peu, ou peu de journaux concentrent la majorité des articles d'un domaine, et beaucoup en concentrent peu.

Ces régularités ont donné lieu à ce que l'on a appelé des lois. Il ne faut pas les comprendre au sens des lois utilisées en physique : elles représentent le phénomène bibliographique, mais ne l'expliquent pas ; ici le terme de loi a donc plutôt le sens de tendance statistique.

Loi de Lotka : elle stipule qu'un très petit nombre d'auteurs publie beaucoup, et une grande majorité d'auteurs publient peu. Par exemple, en observant la distribution du nombre d'articles par auteur, au 21 août 2011, de tous les travaux sur les Amphibiens contenus dans le *Zoological Record*¹, on obtient la courbe de la figure 2.2.

Loi de Bradford : elle stipule qu'un très petit nombre de journaux concentrent la majorité de la littérature d'un domaine, et qu'un grand nombre de journaux en

1. Requête simple ST = "amphibia" pour toutes les années de 1864 à 2011 inclus. N = 72 033 références le 21 août 2011.

FIGURE 2.1 – *Distribution bibliométrique. D'après Rostaing, 1996, p.26.*

contiennent peu. Par exemple, en observant la distribution dans les différents journaux, au 21 août 2011, de tous les articles sur les Amphibiens contenus dans le *ZR*, on obtient la courbe de la figure 2.3.

Loi de Zipf : elle stipule qu'un petit nombre de mots reviennent fréquemment dans un texte littéraire, tandis qu'un grand nombre de mots sont plutôt rares. Zipf avait intitulé sa loi « le principe du moindre effort » car cela suggérait la réflexion que l'être humain choisit et utilise plus facilement des mots familiers que des mots insolites par pure paresse.

Toutes ces lois ont bien sûr été modélisées mathématiquement par leurs auteurs, mais nous n'entrerons pas dans les détails ici. Certains de ces modèles, initialement trop approximatifs, furent améliorés par la suite.

Ces trois lois ont un air de famille qui a poussé certains auteurs à préciser le principe commun qu'elles pourraient partager. Solla Price a le premier tenté de les réunir (Rostaing, 1996, p. 41) avec une théorie reflétant le phénomène connu sous le nom « d'effet Saint-Matthieu », qui veut que l' « on ne prête qu'aux riches » : le succès améliore les chances de succès, tandis que l'échec ne les diminue pas. Cela expliquerait la création d'un pôle cumulatif, et donc par opposition celle d'un pôle de dispersion.

2.2 Indicateurs univariés

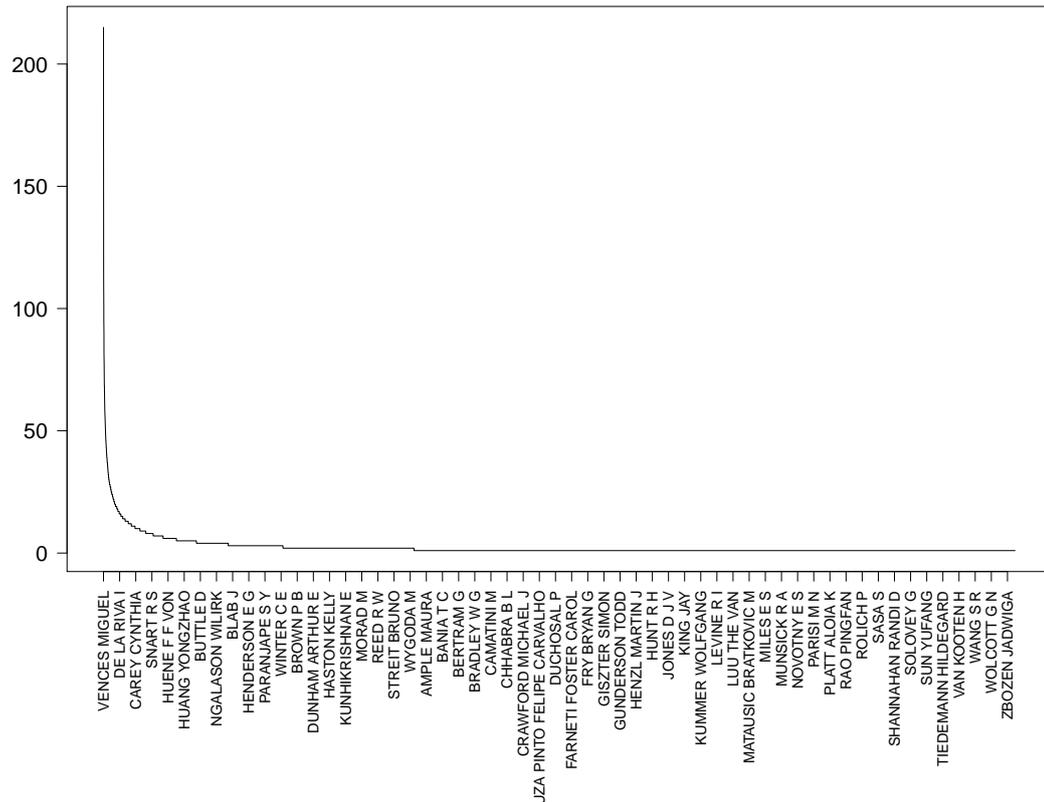


FIGURE 2.2 – Nombre de travaux par auteur pour tous les travaux portant sur les Amphibiens contenus dans le ZR. Recherche réalisée le 21 août 2011 sur tous les travaux publiés jusqu'en 2010. $N = 56\ 473$ auteurs pour 72 033 documents. Indentation de l'abscisse tous les 1 000 auteurs.

Les analyses scientométriques sont traditionnellement classées en deux catégories : celle des indicateurs d'activité et celle des indicateurs relationnels (Callon *et al.*, 1993, p. 39). Elles mettent en jeu deux modèles différents du développement scientifique et technique. Le premier conçoit la science comme une activité productive ordinaire, dont les domaines sont établis, connus et mesurables par des méthodes quantitatives. Le second modèle quant à lui cherche à définir des thèmes de recherche et étudie leurs transformations de manière qualitative.

Ces deux approches sont complémentaires et vont être détaillées ci-après.

2.2 Indicateurs univariés

Les indicateurs d'activité, ou indicateurs univariés, permettent de comparer entre eux les éléments d'un ensemble de références bibliographiques. Ils consistent en deux types de dénombrement (Rostaing, 1996, p. 47) : le comptage des références (mesure de la *productivité*) et le comptage des citations reçues par les références (mesure de l'*impact*).

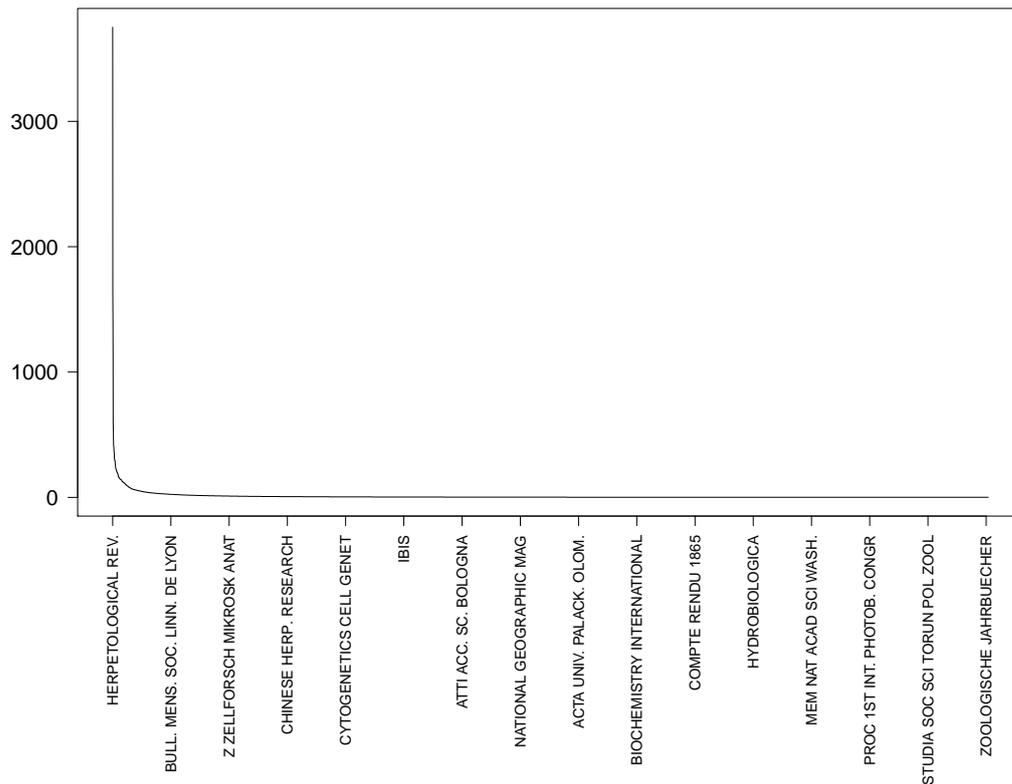


FIGURE 2.3 – Nombre d’articles par journal pour tous les travaux portant sur les Amphibiens contenus dans le ZR. Recherche réalisée le 21 août 2011 sur tous les travaux publiés jusqu’en 2010. $N = 7\,519$ revues pour 69 446 articles. Indentation de l’abscisse toutes les 500 revues.

Les évolutions temporelles de ces indicateurs univariés sont indispensables à connaître pour une analyse plus poussée, de même que l’accélération d’un objet renseigne plus sur sa dynamique que sa vitesse simple.

Ces deux types de dénombrement sont appliqués à la plupart des éléments présents dans une référence bibliographique : auteurs, revues, affiliations, pays, domaines, dates. Ils peuvent être combinés et pondérés pour être améliorés. Toutes les combinaisons de croisements existent et sont potentiellement étudiables, si elles ont un sens et un intérêt (Rostaing, 1996, p. 25).

La nature du calcul permet de distinguer 5 types d’indicateurs (Vinkler, 1988 cité par Rostaing, 1996, p. 58). Ces indicateurs sont présentés tableau 2.1.

Ces mesures font appel à des catégories délimitées a priori. Elles ne donnent aucune indication sur les dépendances et les interactions pouvant exister entre ces catégories : collaborations entre auteurs, pays, disciplines, etc. Ceux-ci sont étudiés par des méthodes qualitatives : les méthodes relationnelles.

La vision positiviste du début de la scientométrie fut conservée par certaines écoles. Pour d’autres, au contraire, l’apparition d’analyses plus qualitatives permit

2.3 Cartes relationnelles de première génération

MESURE	DÉFINITION	EXEMPLE
simple	Comptage simple	Nombre d'articles
spécifique	Productivité en fonction d'un autre facteur	Nombre d'articles par an en fonction du nombre de chercheurs
de balance	Comparaison entre une entrée et une sortie	Nombre de citations données vs nombre de citations reçues
de distribution	Mesure d'une donnée sous forme de part	Nombre d'articles non cités par rapport au total d'articles publiés
relative	Mesure par rapport à une valeur étalon	Nombre de citations par article par rapport à la moyenne du nombre de citations par article dans la discipline

TABLEAU 2.1 – *Les cinq types d'indicateurs bibliométriques définis par Vinkler (1988).*

le développement d'un courant d'étude dont les statistiques ne constituaient pas une fin en soi (Callon *et al.*, 1993, p.6).

2.3 Cartes relationnelles de première génération

Les indicateurs relationnels de première génération n'entrent pas dans le contenu des documents étudiés. Ils font appel à l'analyse des co-publications, l'analyse des réseaux de citation, l'analyse des citations de revues à revues et la méthode des co-citations.

2.3.1 Co-publication

L'analyse des co-publications est réalisée à différents niveaux : auteurs, institutions, pays. Pour chaque élément (une institution, un pays...) on comptabilise ses associations, telles qu'elles sont révélées à travers les adresses mentionnées par les auteurs. En ce qui concerne les pays ou les institutions, chaque publication peut être comptée en *compte de présence* ou bien en *compte fractionné*. Dans le premier cas, elle est comptée autant de fois qu'il y a de pays, ou d'institutions. Dans le second, elle est fractionnée en autant de parts qu'il y a de co-auteurs, de manière à maintenir l'additivité des statistiques (Callon *et al.*, 1993, p. 58). Le nombre obtenu pour chaque association est ensuite ramené au nombre total de publications de l'élément (l'institution, le pays considéré).

2.3.2 Citation de revues à revues, réseaux de citation

La représentation de réseaux de citation de revues à revues est une méthode commode pour obtenir à peu de frais une première représentation d'un champ de recherche. De manière générale, l'étude des réseaux de citations donne une idée de

l'existence de différentes disciplines et des liens qu'elles établissent entre elles. Elle permet également de mettre en relief les réseaux d'influence ou de dépendance entre systèmes nationaux de recherche, ainsi que les jugements portés par la communauté scientifique sur les travaux passés ou en cours (Callon *et al.*, 1993, p. 59-60). Il existe des analyses encore plus fines qui portent sur le contexte de citation. Elles ne sont pas basées sur ce type de méthodologie.

2.3.3 Analyse des co-citations

S'il est difficile d'interpréter une citation prise hors du contexte qui lui a donné naissance, l'apparition simultanée et répétée de deux citations peut être dotée d'une signification plus précise. H. Small, concepteur de cette méthode, émet deux hypothèses (Callon *et al.*, 1993, p. 64).

Si plusieurs chercheurs citent le couple, c'est qu'ils partagent la conviction que les travaux présentés dans ces deux articles sont fortement complémentaires (à la condition que les citations soient distantes au sein de l'article). Chaque auteur cité est considéré par Small comme un symbole, une représentation de tout un ensemble de travaux et de méthodes. Si plusieurs chercheurs citent un couple d'auteurs, c'est donc qu'ils manifestent qu'ils partagent les mêmes représentations collectives de leur activité.

H. Small a établi un lien entre la méthode des co-citations et la notion de paradigme proposée par Kuhn : les documents co-cités sont fondateurs, et les articles qui opèrent ces co-citations contribuent à développer le paradigme dont ils explorent toutes les potentialités (Callon *et al.*, 1993, p. 65).

2.4 Cartes relationnelles de seconde génération

Les indicateurs de seconde génération se donnent à l'inverse pour objet d'analyse les contenus des documents traités (Callon *et al.*, 1993, p. 57). Ils utilisent la méthode des mots associés, basée sur la cooccurrence des mots.

Prenons par exemple l'article *Are Full or Empty Beer Bottles Sturdier and Does Their Fracture-Threshold Suffice to Break the Human Skull?* de Stephan A. Bolliger, Steffen Ross, Lars Oesterhelweg, Michael J. Thali and Beat P. Kneubuehl, paru en avril 2009 dans *Journal of Forensic and Legal Medicine*, vol. 16, n° 3, pp. 138-42. Les mots-clés liés à cet article par les auteurs sont *Breaking energy threshold; Beer bottles; Blunt head trauma*. Cet article établit donc un lien entre trauma crânien, seuil de fracture et bouteilles de bières. On parle de traduction, c'est-à-dire de l'établis-

sement d'une relation d'équivalence entre aires d'action disjointes (Courtil, 1995). Chaque article scientifique crée ces liens entre des notions différentes.

Cette traduction peut être mise en évidence par la méthode des mots associés, développée dans les années 1980 au Centre de sociologie de l'école des mines de Paris (CSEMP) en association avec l'Institut national de l'information scientifique et technique (INIST). Cette notion de traduction est donc un outil précis, permettant d'analyser les rapports entre domaines scientifiques, mais également entre science et société (Courtil, 2006).

Cette approche rejoint donc le courant des *Science Studies*. Si la sociologie des sciences de Merton commença par étudier les acteurs de la science (communautés d'institutions, de chercheurs ; normes et règles les régissant. . .), elle se pencha ensuite sur les conditions de naissance des connaissances scientifiques, s'interrogeant alors sur leur nature tout comme la philosophie des sciences. C'est ce courant qui donna naissance aux jeunes et controversées *science studies*, dont les conclusions relativistes sont très mal reçues par les scientifiques.

La sociologie de Thomas Kuhn, développée par l'ISI, est alors opposée à la sociologie de la traduction développée par le CSEMP (Courtil, 2006, p. 31). « *La recherche ne se réduit pas à l'expérimentation au laboratoire, c'est une activité cognitive et sociale à part entière et cette activité possède, avec la notion de traduction, sa théorie* » (Courtil, 2006, p. 22). L'analyse des contenus fut décriée par l'ISI, qui considérait les mots-clés trop versatiles pour être utilisés.

Je termine cette présentation des techniques de la scientométrie en citant deux de ses auteurs :

the state of the art is 'pre-paradigmatic' : it is an interdisciplinary area integrated only at the level of its subject matter, and an applicational area for various contributing disciplines (Leydesdorff, 2001, p. 3).

Comme toute méthode quantitative, la scientométrie suscite des réactions contradictoires très vives. Les unes sont enthousiastes, les autres sont fortement critiques. Il faut se garder de ces deux excès. La scientométrie rassemble des outils dont il serait absurde de se priver, mais dont il faut connaître les limites (Callon et al., 1993, p. 121).

Chapitre 3

Le *Zoological Record*, base d'étude privilégiée

La base de données bibliographiques la plus exhaustive en zoologie est le *Zoological Record* (ZR), détenue par Thomson Reuters (TR) : « *It covers nearly 6,000 international journals, review annuals, newsletters, popular magazines, monographs, meeting proceedings, books and reports [...] in all major areas of zoology [...] although it is best known for its coverage of systematics, taxonomy, and nomenclature* » (Hughes, 2001). De nombreux auteurs corroborent cette suprématie ; aussi bien des auteurs de travaux indépendants portant sur son contenu et sa couverture (Chisman et Brekke, 1996; Steins, 2001; Parker, 2005; Brown, 2007) ; que les diffuseurs du ZR, concurrents de TR, ou différentes bibliothèques.

Zoological Record Online [®] [...] is the world's most comprehensive index to zoological and animal science literature (Dialog (2011), racheté en 2000 par TR et revendu à ProQuest, un concurrent de TR, en 2008 (Dialog, 2011))

Zoological Record has long been recognized as the "unofficial register" of taxonomic names and systematics, but other topics are also given comprehensive treatment using indexing in both controlled- and natural-language format (ProQuest (2010), un concurrent de TR)

Il s'agit du répertoire bibliographique le plus ancien et le mieux documenté en zoologie et en recherche animale (Bibliothèque de l'université d'Ottawa, 2011)

The Zoological Record covers the worldwide literature on zoology (University of Pittsburgh Library, 2007)

Le *ZR* couvre les articles de zoologie de 1864 jusqu'au présent ; il possède plus de 3 580 000 références bibliographiques provenant de plus de 5 000 journaux ainsi que d'autres sources (livres, rapports, conférences, etc.). Quasiment 72 000 nouveaux enregistrements sont ajoutés chaque année (Thomson Reuters, 2009). C'est donc une source préférentielle pour travailler sur les publications de chercheurs en zoologie et systématique, d'autant qu'elle n'a jusqu'à présent pas été exploitée de manière approfondie dans une perspective scientométrique. Étant donné qu'il n'existe pas de source décrivant de manière très détaillée les caractéristiques du *Zoological Record*, j'ai commencé par compiler et compléter les informations existantes, que je vais vous présenter maintenant.

3.1 Présentation

3.1.1 Historique

Le *ZR* a été créé en 1864 par la Zoological Society of London. Elle le publia annuellement pendant plus de 110 ans, sous le nom de *The Record of Zoological Literature* jusqu'en 1870 puis sous celui de *The Zoological Record* (Bridson, 1968).

En 1977, le *ZR* est produit sous forme informatisée.

En 1980, la *Zoological Society of London* se lie à BIOSIS (organisation à but non lucratif née en 1926, répertoriant les informations en sciences de la vie) afin de mettre fin au délai de trois ans qui s'était instauré avant toute publication du *ZR* (Thorne, 2003). Ils publièrent conjointement le *Zoological Record on CD*[™].

En 1998, BIOSIS lança le *Zoological Record Online*[®] : le *ZR* devint accessible via internet (BIOSIS, 1998).

En 2001, BIOSIS établit un partenariat avec *Cambridge Scientific Abstracts* (CSA, service de résumés en science de la vie à but non lucratif basé au Royaume-Uni), pour produire le *Zoological Record Plus*[®]. Celui-ci inclut alors les résumés de CSA (BIOSIS, 2000).

En 2003, BIOSIS a rendu le *ZR* disponible à de nouveaux vendeurs (NISC, OVID, EBSCO) et enrichi son interface (ajout des champs Broad Terms, Systematics, Super Taxa, Taxa notes) (BIOSIS, 2003).

En 2004, BIOSIS a été acquise par Thomson (Quint, 2003), détenteur jusqu'à peu du monopole du marché de l'information scientifique et technique. Thomson ISI ayant fusionné avec Reuters en 2008 (Euronews, 2007), le *ZR* est aujourd'hui fourni par *Thomson Reuters (TR)*, un géant anglo-canadien du marché de l'information scientifique et technique, listé dans le Stock Exchange (McCook, 2004). « Thomson

3.1 Présentation

Reuters *is the world's leading source of intelligent information for businesses and professionals* [...] *With headquarters in New York and major operations in London and Eagan, Minnesota, Thomson Reuters employs more than 50,000 people in 93 countries* » (Thomson Reuters, 2008). Pour l'anecdote, le groupe Reuters a été créé en 1851 à Londres par « *a German born immigrant specialized in news and stock price information service using a combination of technology including telegraph cables and a fleet of carrier pigeons* » (Thomson Reuters, 2011a) !

3.1.2 Propriétés

Accès au *Zoological Record*

L'accès au *ZR* se fait via une interface graphique dédiée, sur le site de Thomson Reuters ou celui de l'un des différents diffuseurs du *ZR*. L'accès au contenu peut varier suivant les diffuseurs, donc l'interface, choisis (Steins, 2001). L'accès par le MNHN se fait via le portail du Web of Knowledge sur le site de Thomson Reuters (figure 3.1).

The image shows the search interface for the Zoological Record on the ISI Web of Knowledge portal. The page has a green header with the text "ISI Web of KnowledgeSM" and a "DISCOVER" button. Below the header, there are navigation tabs: "All Databases", "Select a Database", "Zoological Record", and "Additional Resources". Under "Select a Database", the "Zoological Record" tab is active. The main search area is titled "Zoological Record®" and contains a "Search for:" section with three input fields. The first field is for "Topic" with an example "bird* migrat* alaska*". The second field is for "Author/Editor" with an example "DaCosta C* OR Da Costa C*". The third field is for "Publication Name" with an example "Journal of Wildlife Management OR Wildlife Research". There are "AND" dropdown menus between the fields and "Add Another Field >>" link. Below the search fields are "Search" and "Clear" buttons, and a note "Searches must be in English". At the bottom, there are "Current Limits" and "Timespan" options. The "Timespan" section has a radio button for "All Years" (selected) and a radio button for "From 1864-1899 to 2011". The "Databases" section shows "Zoological Record --1864-present".

FIGURE 3.1 – Interface de recherche du Zoological Record dans le portail Web of Knowledge de Thomson Reuters [v.4]. Image issue du Zoological Record[®] développé par THOMSON REUTERS[®], Inc. (Thomson[®]), Philadelphia, Pennsylvania, USA : © Copyright THOMSON REUTERS[®] 2010. Tous droits réservés.

La version [5.2] de ce portail, aux possibilités de recherche élargies, est sortie en avril 2011. Les résultats présentés ici ont été obtenus avec l'interface précédente. La nouvelle interface permet d'obtenir les mêmes résultats, à moindre coût (par exemple, la recherche n'est plus limitée aux 100 000 premiers articles).

Outil pour analyses sommaires

Une fois la recherche effectuée, le Web of Knowledge possède un module d'analyse permettant de dénombrer les travaux en fonction d'un critère que l'on spécifie. On peut ainsi compter le nombre d'articles par catégorie, année, organisme étudié, etc.

Sauvegarde des références

L'interface permet de sauvegarder les résultats d'une recherche sous différents formats : format tabulé ('tab'), format balisé ('plain text'), ou encore sous un format spécifique pour trois logiciels de gestion de données bibliographiques : Endnote, RefMan et ProCite. Un exemple du format balisé est présenté figure 3.2.

La sauvegarde via l'interface web ne permet pas de sauver plus de 500 références à la fois.

Champs disponibles pour l'enregistrement

Les données ainsi obtenues contiennent 34 champs pour chaque document. Les principaux champs utilisés dans la suite de ce travail (AN, C1, BD, DE et ST) sont présentés dans le tableau 3.1. L'ensemble des champs est présenté annexe B¹.

Chaque champ peut être rempli avec : des données brutes provenant directement de l'article (ex. : champ « Abstract »), des termes contrôlés issus du thésaurus du ZR (ex. : champ « Broad Terms »), un ensemble des deux (ex. : champ « Systematics »). « *Un thésaurus ou thésaurus de descripteurs, est un type de langage documentaire qui consiste en une liste de termes sur un domaine de connaissances, reliés entre eux par des relations synonymiques, hiérarchiques et associatives. Le thésaurus constitue un vocabulaire normalisé. C'est une sorte de dictionnaire hiérarchisé. Cependant, un thésaurus ne fournit qu'accessoirement des définitions. Les relations des termes et leur sélection l'emportent sur la description des significations* » (Wikipedia, 2011)

1. données issues de http://images.webofknowledge.com/WOKRS53B4/help/ZOOREC/hs_zoorec_fieldtags.html et http://images.webofknowledge.com/WOKRS53B4/help/ZOOREC/hp_full_record.html#dsy1829-TRS_full_record_fields

3.1 Présentation

```
FN ISI Export Format
VR 1.0
PT J
AN ZOOR14706045226
DT Article
TI A review of Yemmatropis (Hemiptera: Lygaeoidea: Berytidae), with
  descriptions of two new species from China.
AU Cai, Bo
  Bu, Wenjun
SO Zootaxa
VL 2808
PS 41–48
PD 4 Apr
PY 2011
UR http://www.mapress.com/zootaxa/list/2011/2808.html
AW http://www.mapress.com/zootaxa/2011/1/zt02808p048.pdf
LA English
ME Online
AB Four species of the genus Yemmatropis Hsiao, 1977, from China are
  treated in this paper, in which Yemmatropis erectus Henry, 2002 is
  newly recorded from China. Yemmatropis hsiaoi sp. nov. and Yemmatropis
  nigrovittatus sp. nov. are described. Photographs of head and pronotum
  and drawings of openings of male genital capsules and parameres of all
  four species, and a key to all Yemmatropis species, are provided to
  assist in identification. The type specimens of the new species are
  deposited in the Institute of Entomology, College of Life Sciences,
  Nankai University, Tianjin, China.
C1 Bu, Wenjun; Institute of Entomology, College of Life Science, Nankai
  University, 94 Weijin Road, Tianjin, 30071, China
EM wenjunbu@nankai.edu.cn
SN 1175–5326; 1175–5334
BD Systematics; Nomenclature; New taxa; Taxonomy; Key; Land zones;
  Palaearctic region; Eurasia; Asia; China
DE Yemmatropis [China / / Distribution; systematics].
  Yemmatropis erectus [China / / New record].
  Yemmatropis hsiaoi; Yemmatropis nigrovittatus [Yunnan / / New
  species].
TN Arthropods; Insects; Invertebrates; True Bugs
ST Animalia, Arthropoda, Insecta, Hemiptera, Heteroptera,
  Pentatomomorpha, Lygaeoidea, Berytidae
OR Yemmatropis – Hsiao 1977 (Berytidae): [Key to species, Extended to
  include new species, p. 48].
  Yemmatropis erectus (Berytidae): [New record].
  Yemmatropis hsiaoi (Berytidae): [Sp nov, Yunnan, P. 42].
  Yemmatropis nigrovittatus (Berytidae): [Sp nov, Yunnan, P. 43].
UT ZOOREC:ZOOR14706045226
ER
EF
```

FIGURE 3.2 – Exemple de référence issue du Zoological Record [v.4] et exportée au format balisé.

Champs disponibles pour la recherche

Le formulaire de recherche du portail Web of Knowledge dédié au ZR permet d'effectuer une recherche dans plusieurs champs de la base. Ils peuvent être interrogés

NOM	DESCRIPTION
AN = <i>Zoological Record</i> Accession Number	A unique identification number for each record in the database
C1 = Author Address	The address of the author who should receive communications about the source publication such as requests for reprints
BD = Broad Terms	Controlled terms from the following hierarchies of the <i>Zoological Record</i> Thesaurus : Subject, Geographical, Palaeontological, Systematic, Taxa Notes. These are broader terms of the controlled terms used to index the source document
DE = Descriptors Data	Controlled and non-controlled terms used to index the source document. Descriptors Data are displayed in a table. Each row of the table contains one or more animal names and related subject index terms. The Organism column contains the name(s) of animals discussed in the source document. The Controlled Term column contains terms from the Subject, Geographical and Palaeontological hierarchies in the <i>Zoological Record</i> Thesaurus. The Modifier column contains free-text or non-controlled indexing terms
ST = Super Taxa	Controlled terms from the Super Taxa hierarchy of the <i>Zoological Record</i> Thesaurus. Super Taxa terms are scientific names displayed to show the taxonomic hierarchy to which animals discussed in the source document belong

TABLEAU 3.1 – *Détail des champs AN, C1, BD, DE et ST du ZR.*

séparément ou simultanément. L'onglet *Advanced Search* permet d'interroger des champs non disponibles dans le mode de recherche simple.

Contrairement à ce que l'on pourrait penser à première vue, les champs de recherche ne correspondent pas tout à fait aux champs mentionnés précédemment, même si les noms sont identiques. Ainsi, le champ « AU » correspond en réalité à une recherche dans « AU » (Author) et « ED » (Editor). Les correspondances sont indiquées dans le tableau 3.2.

À noter :

- deux champs, le champ SD (Subject Descriptors) et le champ TS (Topic), sont disponibles pour lancer une recherche globale sur un grand ensemble de champs ;
- le champ PY doit être combiné avec un autre champ pour que la recherche soit validée.

3.1.3 Informations techniques

J'ai pu obtenir directement la plupart de ces informations auprès de Nigel Robinson, directeur Opérations et Développement du *ZR* en Europe pour *TR* (à York,

3.1 Présentation

RECHERCHE	CHAMPS CONCERNÉS
TI = Title	-
AB = Abstract	-
BT = Broad Terms	-
DE = Descriptors	Descriptors Data table
SY = Systematics/ Organism	Systematics table
BN = Biblio Note	-
TA = Taxa Notes	-
ST = Super Taxa	-
AD = Address	(searches for institution and/or place names)
AU = Author/Editor	Author(s) (both book and journal) Editor(s) fields (searches for author names)
CA = Corporate Author	Corporate Author Book Author fields (within the Author field)
SO = Publication Name	Source field (searches for publication names) The product retrieves books titles and journal titles
PY = Year Published	Source Published field
PU = Publisher	-
IC = Identifying Codes	Accession Number ISSN ISBN
SD = Subject Descriptors	Broad Terms Descriptors Data Super Taxa Systematics Taxa Notes
TS = Topic	Broad Terms Descriptors Data Super Taxa Systematics Taxa Notes Title Book Title Abstract

TABLEAU 3.2 – Champs de recherche disponibles au 10 mars 2010 dans l'onglet de recherche avancée du ZR (Thomson Reuters, 2009). Les champs correspondants du ZR sont donnés dans la colonne de droite.

UK). Il a commencé comme indexeur du ZR pour la Zoological Society of London, puis a continué à travailler pour le ZR, au sein de BIOSIS UK puis de Thomson.

Le thésaurus du ZR

Le thésaurus est constitué de 5 hiérarchies différentes, accessibles en ligne. Les trouver n'est pas évident : il faut cliquer sur la petite loupe en face de Subject Descriptors. Elles sont présentées dans le tableau 3.3.

THESAURUS	DESCRIPTION
Subject Hierarchy	The Subject hierarchy contains controlled terms used to index subject concepts in Zoological Record. It is composed of 39 broad hierarchies
Geographical Hierarchy	The Geographical hierarchy contains controlled terms classified under two hierarchies : Land Zones and Marine Zones
Palaeontological Hierarchy	The Palaeontological hierarchy contains controlled terms classified in the hierarchy Geological Time Periods
Systematic Hierarchy	The Systematic hierarchy contains the Latin/scientific terms used to index animal names. It is composed of 35 broad hierarchies. Do not confuse the Systematic hierarchy, where you find Latin/scientific names, with the Systematics (note the s at the end of Systematics) section of the Subject hierarchy, which contains preferred terms relating to taxonomic changes and nomenclature
Taxa Notes	Taxa Notes terms are broader animal group names referring to the animals discussed in a source document. These are common/vernacular names rather than Latin/scientific names (compare Systematic Thesaurus)

TABLEAU 3.3 – *Les cinq thésaurus du ZR.*

L'ensemble des mots des cinq thésaurus correspond à environ 5 000 termes contrôlés. Les trois premières hiérarchies (Subject Hierarchy, Geographical Hierarchy, Palaeontological Hierarchy) regroupent environ 2 000 termes, et les deux dernières (Systematic Hierarchy, Taxa Notes) environ 3 000 (Nigel Robinson, comm. pers.).

Thesaurus Term:	Add	H	Taxonomy
Scope Note:			The theory and practice of classifying organisms; the relationships of taxa, including positional changes.
Broader term:	Add	H	Systematics
Narrower Term(s):	Add	H	Classification
	Add	H	Included taxa
	Add	H	Key
	Add	H	Numerical taxonomy
	Add	H	Species inquirenda
	Add	H	Status
	Add	H	Taxonomic position
	Add	H	Taxonomic unit
	Add	H	Theoretical taxonomy
	Add	H	Type of classification
Related Term(s):	Add	H	Field identification

FIGURE 3.3 – *Informations contenues dans la division « Subject hierarchy » du thésaurus du ZR pour le mot-clé taxonomy. Image issue du Zoological Record[®] développé par THOMSON REUTERS[®], Inc. (Thomson[®]), Philadelphia, Pennsylvania, USA : © Copyright THOMSON REUTERS[®] 2010. Tous droits réservés.*

On peut voir figure 3.3 que chaque terme est fourni avec :

- sa définition (Scope Note) ;
- ses modifications d'application à travers le temps (Scope Note) ;
- le terme qui le domine dans la hiérarchie (Broader Term) ;

3.1 Présentation

- le ou les termes qu’il domine (Narrower Term) ; et parfois :
- les termes dont il se rapproche, qui sont situés ailleurs dans la hiérarchie (Related Term) ;
- le terme qui est plus approprié si le terme cherché n’est pas le bon.

Les hiérarchisations sont réalisées de deux manières différentes. « *The subject vocabularies [i.e. les 3 premières hiérarchies] are devised by ourselves as a means to organise the published literature. Taxonomic vocabularies [i.e. les 2 dernières hiérarchies] are based on accepted authority sources/texts which we use to build authority files to ensure consistency, accuracy and classification under broader headings within the taxonomic hierarchies. Our generic authority file contains over 250,000 generic level names and we have recorded over 4 million taxonomic names from the literature of the last 145 years* » (Nigel Robinson, comm. pers.).

Le processus d’indexation

Je reproduis ci-contre, en anglais, la description qui m’en a été faite par Nigel Robinson.

« *Currently over 5000 international journals plus monographs and other publications are reviewed by a team of graduate zoologists to select appropriate articles within the scope of ZR. Each article in scope is reviewed and indexed by identification of key concepts and posting those against a series of controlled vocabularies.*

The records in ZR receive broader terms in the hierarchy to allow upposted searching and cascading of narrower terms. Hence all new taxa entries will be upposted to New taxa - Nomenclature - Systematics. Index entries are made using a highly sophisticated electronic data capture system operated by the graduate zoologists. Our computer system is a data entry system linked to a large relational database which prompts our indexers to enter the correct information and checks they have done so correctly. It has over 200 checks to ensure accuracy and consistency and links to our controlled vocabularies of over 3000 terms to a given topic is indexed consistently.

This ensures high levels of accuracy and provides mechanisms to record all taxonomic and nomenclatural changes individually and in a standard format which we can then make available in print or electronically. The views are not those of the individual indexer. We have strict guidelines and policies enforced by a sophisticated computerised data entry system to ensure all indexing is consistent and aligns with our editorial policies and accepted and defined zoological practices as defined by the Code.

Over the years these system have evolved and now have many authority files associated with them together with built in checks and editorial controls to ensure data quality and standardization. This system is separate from the BIOSIS indexing system and employs different principles ».

Il existe également un article sur le sujet, publié en langue chinoise (Hu, 1997).

Particularités des champs documentaires : pièges à éviter

Particularités liées à l'informatisation Jusqu'en 1977, le *ZR* était disponible sous format papier uniquement. En 1977 il fut informatisé et apparut alors sous format numérique. Les articles antérieurs à 1977 furent informatisés a posteriori et compilés dans les *Zoological Record Archive*. « *These new backfiles contain the digitized content of the print Zoological Record Volumes 1-114, 1864-1977. These volumes have been digitized and enhanced with current controlled vocabulary indexing to match the current Zoological Record format. The bibliographic data (authors, titles, source data) found in Archive records are presented largely as published in the original print volumes* » (Thomson Reuters, 2011b).

Il faut donc savoir qu'avant 1977, contrairement aux années 1978 et suivantes, on trouve le champ « SO » (Source) majoritairement disponible sous format abrégé et on ne trouve pas le champ « BN » (ISBN).

Autres particularités Avant 1992, seules les initiales du prénom sont mentionnées. Avant 1994, les documents de plus de dix auteurs contiennent les dix premiers auteurs puis la mention « et al » à partir du onzième ; le champ « C1 » (adresse) n'existe pas.

3.1.4 Distribution des articles dans le *ZR*

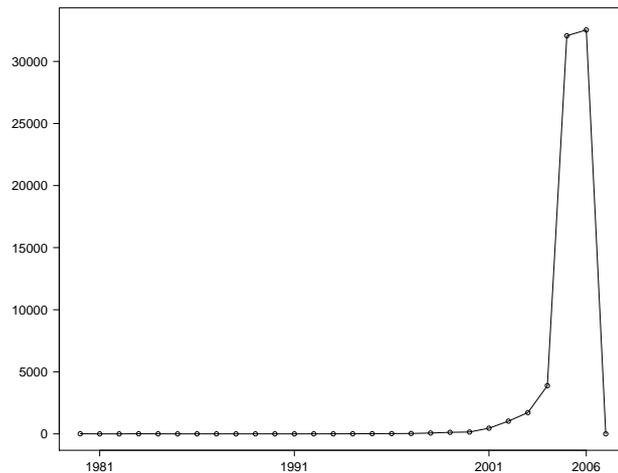
Le nombre total d'articles par année n'est pas une information disponible aux usagers du *ZR*. Il faut donc faire les recherches soi-même.

Chaque article du *ZR* est caractérisé par deux années : l'année de sa publication, d'une part et l'année de son entrée dans le *ZR*, d'autre part. L'année d'entrée dans le *ZR* ne correspond pas toujours à l'année de publication. En effet, si l'on analyse les dates de publication des articles rentrés dans le *ZR* en 2005², on constate qu'ils s'étendent de 1979 à, curieusement, 2006 (figure 3.4a). Il y a quasiment autant d'articles de 2004 que d'articles de 2005 qui ont été informatisés en 2005. Il y a donc un temps de latence non négligeable entre la publication de l'article et son entrée dans le *ZR*. La figure 3.4b montre le nombre de travaux publiés et référencés chaque année dans le *ZR* entre 1980 et 2008. On voit que le nombre de travaux publiés est supérieur au nombre de travaux référencés l'année correspondante ou l'année précédente, excepté de 1995 à 2004 où il semblerait qu'un grand nombre d'articles ait été traité. Le pic de 2000 est particulièrement important.

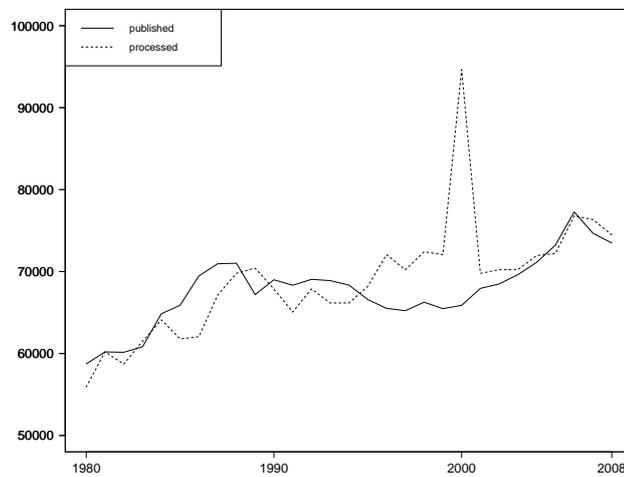
Les contraintes de recherche ne seront pas les mêmes suivant que l'on s'adressera à l'année de publication ou à l'année de référencement (tableau 3.4).

2. Requête IC=ZOO* et Timespan=2005

3.1 Présentation



(a) Date de publication des travaux référencés dans le ZR en 2005. $N = 72\,198$ références



(b) Nombre de références informatisées chaque année dans le ZR depuis 1980

FIGURE 3.4 – Temps de latence entre publication et informatisation dans le ZR.

En outre, le formulaire de recherche n'autorise pas la recherche d'articles par année seulement : il faut impérativement renseigner un autre champ en sus. Or il s'avère que le champ « IC » (Identification Code) n'est jamais nul ; il contient un identifiant pour chaque document indexé dans le ZR, et cet identifiant commence systématiquement par « ZOO ». J'ai donc fait la recherche du mot-clé *ZOO** dans le champ IC afin de couvrir la période couverte par le ZR (1864-2010). J'ai obtenu un total de 3 719 050 articles au 12 mars 2010, ce qui est cohérent avec les 3 580 000 documents annoncés sur le site du ZR (figure 3.5).

INTITULÉ DU CHAMP DOCUMENTAIRE	DESCRIPTION	TRANCHES MAXIMALES DE RECHERCHE
Timespan	Année d'entrée dans le <i>ZR</i>	25 ans pour 1864-1980
Publication year	Année de publication	10 ans

TABLEAU 3.4 – Paramètres de recherche temporelle dans le *ZR*.

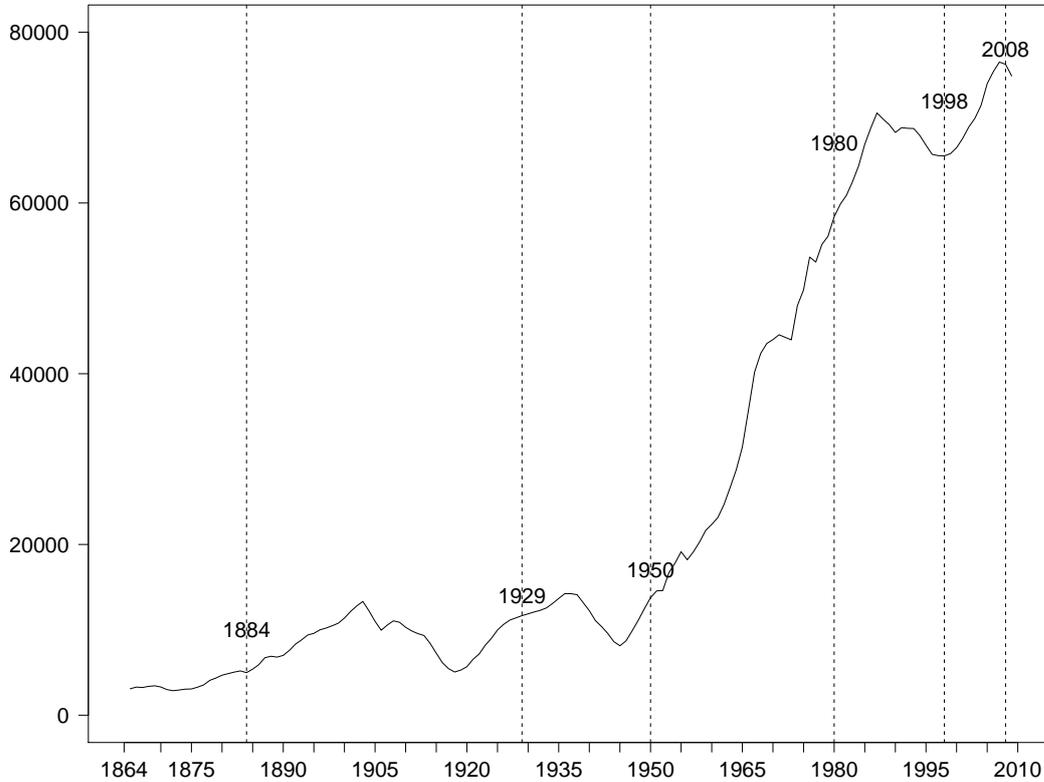


FIGURE 3.5 – Nombre total d'articles dans le *ZR* de 1864 à 2010. Moyenne mobile sur 3 ans (voir définition page 90). $N = 3\ 823\ 175$ références.

On constate que le nombre de travaux augmente significativement après 1945. J'ai déterminé quelles étaient les années qui correspondaient à un milieu de tendance : 1884, 1929, 1950, 1980 ; ainsi que 1998. Elles sont indiquées en pointillés et nous serviront par la suite. Comme expliqué précédemment, j'ai évité de travailler sur des données trop récentes (> 2008) dans la suite de ce travail.

3.1.5 Outils d'analyse du *Zoological Record*

Il n'existe pas d'outil permettant de traiter les données du *ZR*. J'ai donc mis en place un ensemble de fonctions sous **R** (**R Development Core Team, 2011**) permettant de faire les divers traitements utilisés dans la suite de ce travail. Deux exemples de scripts sont donnés en annexe C : le script automatisé d'importation des notices, et le script de déduplication des articles.

3.2 Construction d'un corpus de travail issu du ZR

Pour écrire et tester ces fonctions j'ai travaillé sur un petit sous-ensemble cohérent de références. J'ai choisi comme sous-ensemble celui des travaux portant sur une famille d'amphibiens, les Pelobatidae, pour le petit nombre qu'il représente (2 328 références) et pour sa diversité thématique. La construction de ce corpus est présentée dans le chapitre suivant.

L'une des difficultés les plus importantes en scientométrie, et peut-être la plus gourmande en temps, est celle de la détection/correction des erreurs de données. Le dernier chapitre de cette partie présente donc en détail la méthodologie suivie pour détecter les duplicata des noms d'auteurs.

3.2 Construction d'un corpus de travail issu du ZR

Si l'on fait une recherche dans une base de données bibliographiques dans le but de recueillir le corpus d'articles d'un thème donné, on doit faire particulièrement attention aux mots utilisés pour réaliser la recherche. La construction d'une requête de recherche est l'étape critique de la constitution d'un corpus à partir d'une base de données. Cette construction nécessite des aller-retours incessants entre les mots-clés utilisés, les champs de recherche choisis et les résultats obtenus. Elle nécessite de connaître les méthodes d'indexation des bases utilisées, ainsi que leurs limites.

En effet, deux écueils peuvent se présenter :

- que la recherche soit trop large, et renvoie à des articles qui sont sans rapport avec le sujet recherché ; on les appelle des faux positifs, ou **bruit** ;
- que la recherche soit trop restreinte, et ne renvoie pas une partie des articles concernant le sujet recherché ; on les appelle des faux négatifs, ou **silences**.

Toute la construction des expressions de recherche doit donc chercher à minimiser ces deux types d'erreur. La partie suivante explique la démarche suivie pour construire ce corpus.

Dans notre cas, ces expressions de recherche peuvent être modulées de différentes façons :

- en choisissant de lister soit les genres, soit les familles concernées ;
- en choisissant de rechercher avec le nom latin ou le nom vernaculaire ;
- en présentant ces noms soit sous forme entière, soit sous forme tronquée.

Chacun de ces types de recherche a une influence sur le type d'erreur minimisé.

3.2.1 Faut-il chercher par genres ou par familles ?

Afin de répondre à cette question j'ai réalisé plusieurs tests de recherche en utilisant le nom de genre *Pelodytes* et le nom de famille Pelodytidae. Comme il s'agissait

de savoir si tous les articles indexés avec un nom de genre l'étaient également avec un nom de famille, et vice versa, j'ai choisi de faire les tests sur une famille monogénérique (tableau 3.5).

N°	Recherche	Nombre d'articles	%
1	TS = pelodytes OR TS = pelodytidae	210	100
2	TS = pelodytes	194	92
3	TS = pelodytidae	87	41
4	TS = pelodytidae NOT TS = pelodytes	16	18
5	TS = pelodytes NOT TS = pelodytidae	123	63

TABLEAU 3.5 – *Recherches effectuées dans le ZR pour construire le corpus de références sur les Pelobatidae - partie I.*

Une recherche avec le nom de genre (2) amène plus de résultats qu'une recherche avec le nom de famille (3). 18 % des articles avec le nom de famille ne contiennent pas le nom de genre (4). Il s'agit principalement d'articles portant sur plusieurs familles (révisions, systèmes, faunes, phylogénies, anatomie), mais également d'articles avec faute de frappe (*Pelodytes punctatus*). 63 % des articles avec le nom de genre ne contiennent pas le nom de famille (5). Il s'agit aussi bien d'articles traitant du niveau spécifique que d'articles traitant de niveaux plus élevés comme la famille.

Ces résultats montrent que l'indexation des articles par des mots-clés de nature nomenclaturale est loin d'être réalisée puisque la hiérarchie des taxons n'est pas présente de façon exhaustive dans l'indexation. Une recherche portant uniquement sur les genres ou uniquement sur les familles génère à chaque fois du silence. Il faut donc utiliser à la fois les mots-clés de genre et les mots-clés de famille pour effectuer les recherches.

3.2.2 Faut-il faire une recherche avec des termes complets ou utiliser les troncatures ?

La recherche par termes exacts nécessite de pouvoir lister tous les noms existant pour le taxon recherché. On pourrait donc penser, à première vue, qu'il est plus pertinent de faire une recherche avec troncature ; à la fois pour réunir tous les termes dans une recherche courte (ex. *Scaphiop** pour rechercher *Scaphiopus*, Scaphiopodidae), et pour rassembler le plus de documents possibles (prenant en compte les noms anglophones, issus d'une légère modification du nom latin, par ex. *Scaphiopodid* pour Scaphiopodidae).

Mais cela pose alors la question du bruit, c'est-à-dire le renvoi d'articles portant sur des taxons qui ne sont pas en relation avec les taxons recherchés. Classiquement, ces renvois peuvent en effet avoir lieu suite à :

3.2 Construction d'un corpus de travail issu du ZR

- des erreurs de frappe (on obtient alors des articles traitant du Colubridae *Scaphiophis hurterii* au lieu du Scaphiopodidae *Scaphiopus hurterii*);
- des homonymies (ex. : le parasite *Diplorchis scaphiopus*; le coléoptère *Afroplectus scutigera* ou encore l'échinoderme *Cidarites scutigera*).

Or, le bruit est également renforcé par l'utilisation de troncatures : une troncature trop précoce peut produire des résultats très larges (ex. : *scaphiop** qui renvoie des articles traitant du genre d'amphibiens *Scaphiophryne*).

Le tableau 3.6 montre l'accroissement du nombre de taxons étrangers retournés (avant-dernière colonne) et du nombre de documents (dernière colonne) avec l'usage des troncatures.

Recherche exacte			Recherche tronquée			Croissance	
Expression	NT	ND	Expression	NT	ND	NT (%)	ND (%)
<i>pelobatidae</i>	0	1963	<i>pelobatid*</i>	2	1984		1
<i>pelodytidae</i>	0	87	<i>pelodytid*</i>	0	88		1
<i>megophryidae</i>	0	61	<i>megophryid*</i>	0	66		8
<i>scaphiopodidae</i>	0	12	<i>scaphiopodid*</i>	0	12		0
<i>borneophrys</i>	0	0	<i>borneophr*</i>	0	0		
<i>brachytarsophrys</i>	0	17	<i>brachytarsophr*</i>	0	17		0
<i>leptobranchium</i>	1	90	<i>leptobranch*</i>	22	114	2100	27
<i>leptolalax</i>	0	46	<i>leptolax*</i>	184	216		370
<i>megophrys</i>	0	152	<i>megophr*</i>	0	212		40
<i>ophryophryne</i>	0	19	<i>ophryophry*</i>	0	19		0
<i>oreolalax</i>	0	46	<i>oreolax*</i>	11	57		24
<i>scutigera</i>	24	122	<i>scutig*</i>	200	764	733	526
<i>xenophrys</i>	0	20	<i>xenophr*</i>	3	23		15
<i>pelobates</i>	0	821	<i>pelobat*</i>	15	2186		166
<i>pelodytes</i>	0	194	<i>pelodyt*</i>	8	228		17
<i>scaphiopus</i>	1	730	<i>scaphiop*</i>	41	836	4000	14
<i>spea</i>	0	172	<i>spea</i>	0	172		0

TABLEAU 3.6 – Recherches effectuées dans le ZR pour construire le corpus de références sur les Pelobatidae - partie II. Recherche au 1^{er} décembre 2010. NT : nombre de taxons étrangers trouvés dans l'Index of Organisms Names. Rq : on compte les taxons, pas les formes syntaxiques. Ainsi *Aedes* (*Aedimorphus*) *leptolabis* et *Aedimorphus* *leptolabis* comptent pour un seul taxon étranger. *Megophryidae* n'est pas compté comme étranger à *Megophrys*. ND : nombre de documents trouvés dans le ZR. Croissance NT = $\frac{NT_{tronquée} - NT_{exacte}}{NT_{exacte}} \times 100$. Croissance ND = $\frac{ND_{tronquée} - ND_{exacte}}{ND_{exacte}} \times 100$.

On remarque que cet accroissement est variable suivant les taxons recherchés : l'accroissement du nombre de taxons étrangers peut aller jusqu'à 2100 %, et l'accroissement du nombre de documents peut aller de 1 % à 526 %.

Trois moyens peuvent être envisagés pour éliminer ces faux positifs :

- ajouter dans la recherche un mot-clé supposé présent pour chaque article sur les Pelobatidae (ex. : *amphibia*);
- éliminer les articles traitant de taxons étrangers au taxon recherché;
- ne pas utiliser les troncatures.

N°	Recherche	Nombre d'articles	%
1	TS = pelodytes	194	100
2	TS = pelodytes NOT TS = amphibia	2	1
3	TS = scaphiopus	730	100
4	TS = scaphiopus NOT TS = amphibia	15	2
5	TS = scutigera	122	100
6	TS = scutigera NOT TS = amphibia	42	34

TABLEAU 3.7 – *Recherches effectuées dans le ZR pour construire le corpus de références sur les Pelobatidae - partie III.*

Si l'on réalise des recherches en appliquant la première méthode (tableau 3.7), on constate qu'entre 1 et 34 % des articles indexés par le nom de genre recherché ne possèdent pas le mot-clé *amphibia*. Prenons l'exemple des *Scaphiopus* : recherches (3) et (4). Parmi les articles laissés de côté, certains représentent effectivement du bruit (articles qui ne traitent pas d'amphibiens) et sont donc repérés comme on le souhaitait par la méthode (1 article sur 15). Cependant, d'autres sont au contraire en lien avec le sujet recherché (14/15) mais n'ont pas été indexés avec *amphibia* à la suite :

- d'erreurs d'indexation (ex. : Colubridae au lieu de Pelobatidae ; Insecta au lieu d'Amphibia) ;
- une indexation incomplète (ex. : Reptilia au lieu de Reptilia et Amphibia ; articles portant sur des parasites, des prédateurs ou des proies du taxon étudié).

Si on prend maintenant l'exemple du taxon *Scutigera* (recherches (5) et (6)) on voit que le pourcentage d'articles éliminés est beaucoup plus important. Cette fois, tous les articles correspondent à du bruit (42/42). En effet il existe plus de 20 taxons homonymes, dans divers embranchements (coléoptère *Afroplectus scutigera*, échinoderme *Cidarites scutigera*...)

Au final, le pourcentage d'articles éliminés par cette méthode alors qu'ils devraient être conservés est faible.

Je propose donc de faire une recherche avec des termes exacts de genre et de famille pour limiter le bruit généré par les troncatures, et le mot-clé *amphibia* pour supprimer ensuite du corpus les articles traitant des taxons bruyants.

On obtient alors 2 373 références. J'ai utilisé l'outil *Analyze results* pour obtenir la liste exhaustive des organismes traités dans ces articles. J'ai retenu les noms qui totalisaient au moins 12 articles (soit 5 % du total) et l'ai transmise à Annemarie Ohler (MNHN) pour identifier d'éventuels intrus. J'ai ensuite vérifié que ces taxons étaient bien liés d'une façon ou d'une autre, avec la famille étudiée (parasites, proies...).

3.3 Gestion des auteurs synonymes

3.2.3 Faut-il rechercher par les noms vernaculaires ?

À première vue, une recherche par nom vernaculaire paraît contre-indiquée. Leurs délimitations sont le plus souvent vagues, voire chevauchantes. Cependant, il est pertinent de se demander si des travaux de biochimie, par exemple, vont prendre la peine d'indiquer précisément dans leur titre, mots-clés ou résumé le nom latin de l'organisme étudié. J'ai donc fait une recherche en ajoutant le mot *spadefoot** à la recherche précédente. On obtient 88 articles en plus, soit 3,7 % du total. Étant donné le nombre important de noms vernaculaires existant (Frost, 2011), et le faible nombre d'articles concernés par ce silence, je choisis de ne pas faire de recherche par nom vernaculaire dans le *ZR*.

3.2.4 Requête finale

La famille des Pelobatidae comprend aujourd'hui quatre espèces du même genre *Pelobates*, distribuées à l'ouest de l'Eurasie et au nord-ouest de l'Afrique. Avant 1973 elle comportait également :

- le genre d'Eurasie *Pelodytes* qui forme aujourd'hui la famille des Pelodytidae.

Avant 1981 elle comportait également :

- les genres d'Amérique du Nord *Scaphiopus* (3 espèces) et *Spea* (4 espèces); qui constituent aujourd'hui la famille des Scaphiopodidae.

Avant 1993 elle comportait également :

- les onze genres et 150 espèces d'Asie qui forment aujourd'hui la famille des Megophryidae.

Quelques photos de Pelobatidae sont présentées figure 3.6.

Requête finale (TS = Topic) :

TS = ("pelobatidae" or "pelobates" or "megophryidae" or "pelodytidae" or "scaphiopodidae" or "borneophrys" or "brachytarsophrys" or "leptobranchella" or "leptobranchium" or "leptolalax" or "megophrys" or "ophryophryne" or "oreolalax" or "scutigera" or "xenophrys" or "pelobates" or "pelodytes" or "scaphiopus" or "spea")

3.3 Gestion des auteurs synonymes

Les données de co-publication issues des bases de données bibliographiques se présentent sous forme textuelle. Un champ typique se présente sous la forme suivante :

Mollov, Ivelin A. ; Velcheva, Iliana G.



Pelobates syriacus

©Horia Bogdan



Scaphiopus couchii

©2004 William Flaxington



Megophrys nasuta

©2006 Wouter Beukema



Pelodytes punctatus

©2006 Iñigo Martínez-Solano

FIGURE 3.6 – Photographies de Pelobatidae.

Chaque collaboration est présentée par une chaîne de caractères contenant une succession d'auteurs, présentés suivant l'ordre de l'article et séparés par un caractère délimiteur. Or, plusieurs noms peuvent exister pour une même personne, et un même nom peut faire référence à différentes personnes. Sur une échelle des dissemblances (figure 3.7), on peut dire que les erreurs d'identification peuvent aller de l'existence de pseudonymes (deux noms différents pour une même personne) à l'existence d'homonymes (un même nom pour deux personnes), en passant par l'existence d'erreurs typographiques, de variantes orthographiques issues de multiples translittérations, de modifications de l'encodage des noms dans la base de données, ou encore de changements civils de nom. Dubois (2003a) en donne quelques exemples dans le cadre des publications zoologiques.

Je me suis intéressée à réduire le nombre d'erreurs de la première série : celles qui consistent en l'existence de plusieurs noms pour une même personne physique.

J'ai donc laissé de côté le problème des homonymes (un même nom pour deux personnes), qui nécessiterait de vérifier la cohérence du profil professionnel (spécialité,

3.3 Gestion des auteurs synonymes

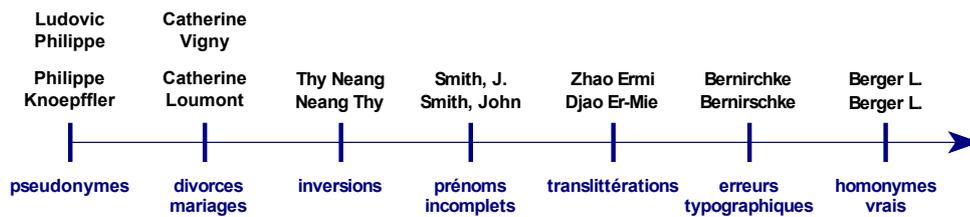


FIGURE 3.7 – Les différentes erreurs que l'on peut retrouver dans la dénomination des auteurs.

pays...) de tous les auteurs ayant écrit plus d'un article. Ces homonymes peuvent être détectés a posteriori, lorsque les résultats obtenus présentent des incohérences.

3.3.1 Quelle méthode utiliser ?

L'optimisation de la détection de données dupliquées est une thématique active de recherche qui fait suite à l'accroissement des bases de données et à l'automatisation des tâches effectuées sur ces bases. Le besoin de pouvoir garantir la fiabilité de sources agrégées, ou de très grand volume, est ainsi un souci central et transversal. Le processus qui en résulte porte donc des noms différents suivant les communautés qui le développent : le terme *record linkage* (couplage d'enregistrement) des épidémiologistes et des statisticiens devient ainsi le *data matching* ou l'*object identity problem* des informaticiens, quand ce n'est pas... *object identification*, *datacleaning*, *approximate matching* ou *approximate joins*, *fuzzy matching* ou encore *entity resolution* (Winkler, 2006).

De nombreuses méthodes ont été déployées depuis les années 1950 pour optimiser cette détection de doublons. Dans sa revue du domaine, Winkler identifie que pas moins de 13 pistes de recherche différentes sont explorées aujourd'hui (Winkler, 2006). La plupart de ces pistes de recherche restent néanmoins basées sur le formalisme initial, probabilistique, de Fellegi et Sunter (1969). De façon simplifiée, la méthode est la suivante : si l'on considère quatre enregistrements contenant chacun cinq champs (par exemple : nom, prénom, date de naissance, lieu de naissance, boisson préférée), chaque paire possible d'enregistrement se verra comparée, pour chacun de ces cinq champs, et un poids final issu de ces différentes comparaisons lui sera attribué. L'ensemble de ces poids est ensuite utilisé pour classer les paires dans différentes catégories : liées, non liées, ou probablement liées. Dans ce dernier cas, l'intervention d'un utilisateur sera nécessaire pour trancher ; par exemple à l'aide de données supplémentaires. Les méthodes mentionnées précédemment varient dans les fonctions de comparaison, de classification ou d'optimisation qu'elles mettent en œuvre.

Dans notre cas, nous n'avons à notre disposition qu'un nombre limité de champs : celui du nom d'auteur, éventuellement séparé en deux champs distincts « nom de famille » et « prénom ». Il est en effet impossible d'obtenir des valeurs d'attributs simples pour chaque auteur à partir de ces bases de données bibliographiques : chaque auteur est susceptible d'avoir plusieurs années de publication, sujets, taxons étudiés ou encore laboratoires. Les méthodes probabilistiques citées ci-dessus ne sont donc pas adaptées (Murat Sariyar, com. personnelle). On fait donc appel dans ce cas à des mesures de distance entre chaînes textuelles et on choisit de considérer que deux noms sont potentiellement synonymes lorsque cette distance est supérieure à un seuil défini. Un exemple de distance est celui de la distance de Levenshtein : elle est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à l'autre.

3.3.2 Outils à disposition

Aucun logiciel d'analyse en scientométrie disponible en libre accès ne propose une option de déduplication des données satisfaisante. Elle est rarement implémentée (Cobo *et al.*, 2011), ou, lorsque c'est le cas, c'est sous une forme excessivement simple puisque la ségrégation des paires est réalisée de manière automatique : en se basant sur le seuil de distance précédemment décrit et défini par l'utilisateur, le logiciel tranche purement et simplement sans donner la possibilité de recourir au contexte de travail de chaque auteur. L'utilisateur n'a donc pas la possibilité de tenir compte du contexte pour valider chaque synonyme potentiel. J'ai donc mis en place une procédure nouvelle pour traiter les synonymies des données d'auteurs, processus dont j'ai eu besoin par la suite.

L'ensemble des résultats qui vont suivre, obtenus lors de la mise en place du processus, proviennent du corpus précédemment défini, portant sur les Pelobatidae.

3.3.3 Algorithme d'homogénéisation

Principe

Comme vu précédemment, un champ typique de collaboration se présente sous la forme suivante :

Mollov, Ivelin A. ; Velcheva, Iliana G.

Le principe de la démarche suivie pour homogénéiser les auteurs est donc le suivant :

1. normalisation des données ;

3.3 Gestion des auteurs synonymes

2. extraction des auteurs ;
3. recherche des inversions « Nom - Prénom » ;
4. calcul de la distance entre toutes les paires d'auteurs ;
5. vérification manuelle (*clerical review*) de chacune des paires en tenant compte des informations de publication pour chacun des deux auteurs : titres de leurs écrits, co-auteurs, et éventuellement années de publication ;
6. remplacement de chaque nom considéré comme synonyme. Si plusieurs synonymes existent pour une même personne, on conserve le synonyme le plus long.

Il n'existe pas de fonction dédiée à ce type de travail sous R. La librairie la plus approchante, *RecordLinkage* (Sariyar et Borg, 2010), ne prend en charge que l'étape de calcul des distances mais rien n'est ensuite prévu pour la validation des paires en tenant compte du contexte ou le remplacement final des données : « *we are aware that RecordLinkage lacks an important subtask of record linkage, the standardization of records. However, this is an area that itself would need extensive research efforts [...] it is therefore left to the users to tailor standardization methods fitting the requirements of their data* » (Sariyar et Borg, 2010). Il m'a donc fallu coder la plus grande partie de l'algorithme.

Difficultés rencontrées

Limites d'un travail sur la similarité des noms Les cas d'utilisation de pseudonymes ou de changements civils de nom ne peuvent être mis en évidence au regard de la différence qu'ils peuvent entraîner. Ils n'ont donc pas pu être traités ici. Au regard de leur diversité, (ONU, 2011), les variantes de translittération peuvent également difficilement être détectées.

J'ai donc traité uniquement les inversions, les erreurs typographiques et les noms présents sous forme abrégée (avant 1992, le prénom était rentré sous forme d'initiale dans le *Zoological Record*).

Normalisation des données. La normalisation des données a été une étape délicate, consistant à repérer et corriger les erreurs de ponctuation, supprimer la ponctuation superflue, mettre tous les noms en majuscules afin de s'affranchir de la casse.

Dans ces données de co-publication, deux noms d'auteurs sont délimités par un point-virgule tandis que le nom et le prénom d'un même auteur sont délimités par une virgule :

Nom 1, Prénom 1 ; Nom 2, Prénom 2

Certaines erreurs de typographie étaient présentes au niveau de la ponctuation : les points-virgules pouvaient être remplacés par « ., » et la virgule pouvait manquer. On obtenait ainsi

Mollov, Ivelin A., Velcheva, Iliana G.
Abrol Bhavna ; Achim, A.

Une première étape a donc consisté à remplacer « ., » par « ; » (2 articles sur 2321). Notons que ce remplacement n'a pas été effectué lorsque « ., » était suivi de « Jr », « jun » ou « II » (30 articles sur 2321). Une seconde étape a consisté à ajouter une virgule après le premier groupe de termes d'un nom lorsque celle-ci manquait (57 noms sur 2779).

En outre, je me suis rendue compte que la présence de points ou de tirets portait à conséquence sur le calcul de la similarité. Ainsi, la similarité entre

« Shuqiang » et « S. »

est de 0,59 alors que la similarité entre

« Shuqiang » et « S »

est de 0,74. J'ai donc choisi de supprimer toute la ponctuation superflue à l'exclusion des virgules et point-virgules. L'ensemble des noms a finalement été mis en majuscule pour s'affranchir de la casse lors de la comparaison. Au final,

Mollov, I.-A., Velcheva Iliana G.

devient

MOLLOV, I A ; VELCHEVA, ILIANA G

Cas des auteurs inversés. Dans notre échantillon, huit auteurs étaient présents à la fois sous une forme « Nom + Prénom » et « Prénom + Nom ». Pour les détecter, j'ai mis en place une procédure de recherche déconstruisant chaque nom sous sa forme « Nom + Prénom », inversant les deux termes, puis recherchant cette nouvelle forme dans l'ensemble des noms. L'existence d'auteurs ayant un prénom sous forme abrégée m'a conduite à effectuer cette recherche sous la forme « Nom + Initiale ». Ainsi, les synonymies du type

HOI-SEN, Y. et YONG HOI-SEN

ont pu être mises en évidence.

Cas des erreurs typographiques et des variantes d'encodage.

Choix de la métrique de similarité. La librairie *RecordLinkage* propose, au choix, deux métriques de similarité de chaînes de caractères : la mesure de similarité de Levenshtein et la mesure de similarité de Jarowinkler. Elles prennent des valeurs allant de 0 à 1 (0 : chaînes très dissemblables, 1 : chaînes identiques).

La mesure de similarité de Jarowinkler étant particulièrement recommandée pour traiter des chaînes de noms, c'est celle que j'ai retenue. Ainsi, les trois paires :

« COLUCCI MICHEL GÉRARD JOSEPH » et « COLUCHE MICHEL GÉRARD
JOSEPH »
« COLUCCI » et « COLUCHE »
« MICHEL GÉRARD JOSEPH » et « M G J »

auront respectivement des similarités de 0,91 ; 0,89 et 0,62.

Comme on vient de le voir avec cet exemple, la similarité peut être détectée de plusieurs manières différentes : en s'intéressant au nom complet, au nom de famille, au prénom, ou encore à une combinaison des trois. Dans ce contexte, quelle chaîne de caractère utiliser ?

Choix de la chaîne textuelle à comparer. La figure 3.8 présente la répartition des paires d'auteurs en fonction de la mesure de similarité de jarowinkler pour les trois types de chaînes : nom complet, nom de famille et prénom. Cette figure nous donne deux informations. La première est que, quelle que soit la chaîne textuelle sur laquelle on fait la comparaison, il n'existe pas de valeur pour laquelle on observerait un « décrochage » net, donc pas de valeur permettant de définir clairement un seuil. La seconde information concerne le nombre de paires à vérifier pour chacun des seuils potentiels. On voit ainsi que pour le choix d'un seuil de 0,95, il y aurait 13 005 paires à vérifier dans le cas des prénoms, 1 241 dans le cas des noms de famille et 186 dans le troisième cas. Ces nombres augmentent très rapidement lorsque l'on change de seuil : ils passent respectivement à 14 506, 2006 et 595 pour un seuil de 0,90. S'il est envisageable de vérifier quelques centaines de paires, il est évident que vérifier un millier ou plusieurs milliers de paires ne fait aucun sens. Leur nombre se doit donc d'être diminué au maximum sans que cela soit fait au détriment de l'information recherchée.

Prenons le premier cas, celui où l'on comparerait seulement les noms complets. Un problème rencontré est celui de la grande différence qui peut exister dans les formes de prénom pour un même auteur ; notamment entre la forme abrégée « Nom + Initiales »

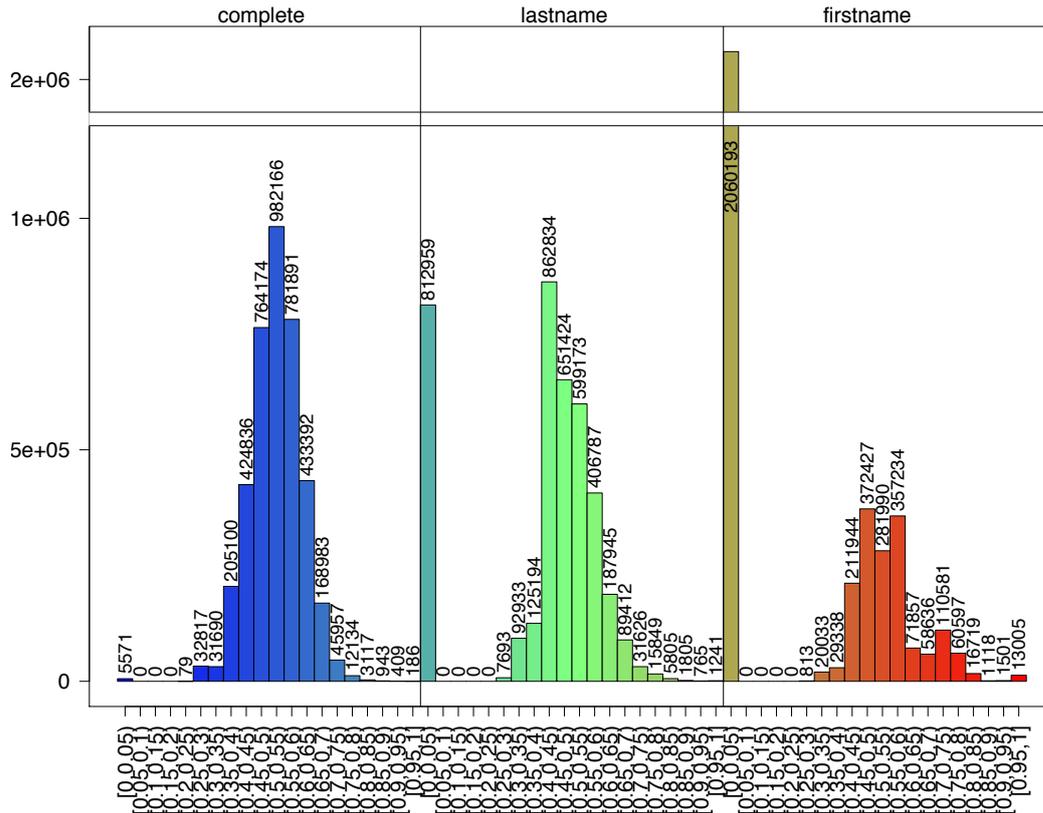


FIGURE 3.8 – Répartition des paires d’auteurs en fonction de la mesure de similarité de Jarowinkler pour trois types de chaînes : nom de famille et prénom, nom de famille seul, prénom seul. 0 : entièrement dissemblables, 1 : identiques. $N = 3\,913\,003$ paires.

qui prévaut avant 1992 et la forme complète « Nom + Prénom » qui prévaut après 1992. Ainsi, la paire :

LIUKIN S Q - LIUKIN SHUQIANG

a une similarité de 0,90 alors qu’ils s’agit de la même personne. Ce phénomène est renforcé pour des noms courts. Ainsi, la paire :

LI S Q - LI SHUQIANG

a une similarité de 0,85. Or, ces seuils risquent d’amener de nombreux faux positifs. Par exemple, la paire :

ANDERSON JIM - ANDERSON JOHN

qui fait clairement référence à des personnes différentes, a une similarité de 0,92.

La figure 3.9 nous montre la répartition des prénoms en fonction de leur longueur. On voit que les prénoms abrégés (1 ou 2 lettres) sont en effet légion et que ce problème

3.3 Gestion des auteurs synonymes

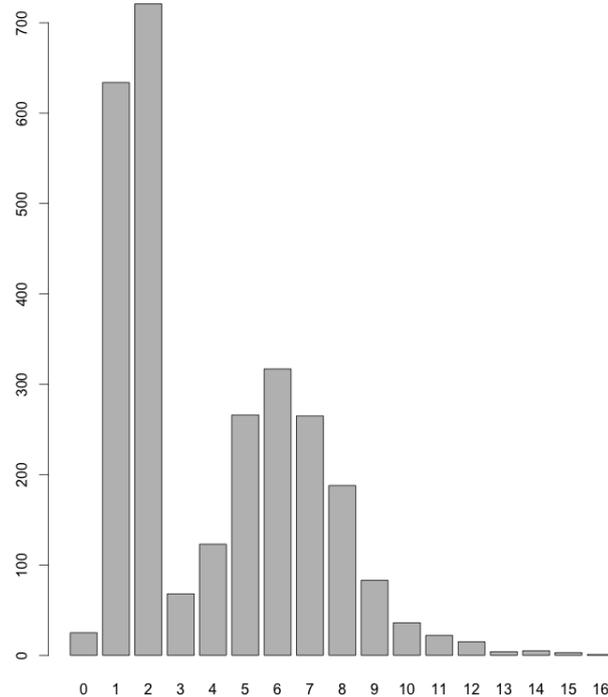


FIGURE 3.9 – Répartition des prénoms en fonction de leur longueur. Les espaces présents ont été supprimés. Le même prénom peut apparaître deux fois.

doit donc être pris en compte. La solution semble donc de faire une recherche couplée sur les chaînes de prénom et les chaînes du nom de famille, en imposant deux seuils. Il reste alors à répondre à la question : quel(s) seuil(s) choisir ?

Choix des seuils. Le choix du seuil est critique car il conditionne le risque de considérer comme non-synonymes des paires qui sont en réalité synonymes.

Regardons tout d’abord le choix du seuil pour les prénoms. La figure 3.9 nous indique que la longueur maximale obtenue pour un prénom est de 16 lettres (« TIBERIU CONSTANTIN »). Par conséquent, je propose de fixer le seuil de similarité des prénoms à la valeur de mesure de similarité entre les deux chaînes

« S » et « SAAAAAAAAAAAAAAAAA »

(la dernière ayant une longueur de 16 lettres). Ainsi, tous les prénoms abrégés devraient pouvoir être repérés. Cette mesure est 0,71875.

Intéressons-nous maintenant au seuil pour les noms de famille.

La figure 3.10-A nous montre le nombre de paires obtenues avec le choix d’un seuil de 0,71875 pour les prénoms et différents seuils pour les noms de famille (de

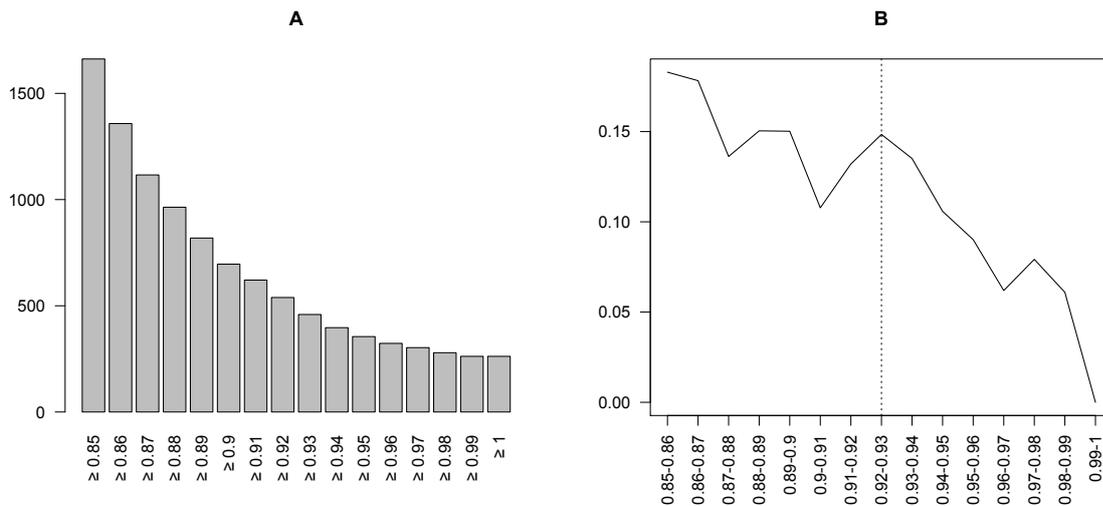


FIGURE 3.10 – A : nombre de paires obtenues en faisant varier le seuil de similarité des noms de famille, pour une similarité de prénom supérieure ou égale à 0,71875. B : taux d'accroissement du nombre de paires entre les différents seuils.

0,85 à 1). La figure 3.10-B nous montre le taux d'accroissement du nombre de paires entre les différents seuils. On constate, d'une part, que le seuil de 0,92 correspond à une modification dans le profil du taux d'accroissement, et d'autre part, que c'est un seuil pour lequel le nombre de paires à vérifier n'est pas trop élevé (350).

Je propose donc de vérifier toutes les paires pour lesquelles la similarité de nom de famille est supérieure ou égale à 0,92 et la similarité de prénom supérieure à 0,71875, soit 350 paires.

Module de vérification. Cette étape, qui demande une validation de l'utilisateur pour chaque paire vérifiée, nécessitait la mise en place d'un processus interactif entre R et l'usager. Pour ce faire j'ai utilisé les outils de la librairie *tcltk*, spécifiquement dédiée à la création de fenêtres d'interaction.

Reconstruction des listes de synonymes et choix de la forme à conserver. Cette étape a demandé l'élaboration d'une fonction supplémentaire afin de reconstruire l'ensemble des synonymies et de conserver le nom le plus long parmi tous les synonymes trouvés pour un auteur.

Les outils étant en place, intéressons-nous maintenant de plus près à la taxonomie.

Deuxième partie

La taxonomie, une discipline isolée

Chapitre 4

Une discipline aux contours flous

*the definition of taxonomy differs depending upon
your own personal or scientific perspective
(Knapp, 2007)*

4.1 Étymologie

4.1.1 Création

Le terme taxonomie est créé par Candolle en 1813. Le terme est donné une première fois page 19 : la taxonomie, « *mot formé de τὰξις ordre, et νόμος loi, règle* », est la « *théorie des classifications appliquée au règne végétal* ». Il est ensuite défini p. 25-26 : « [...] *nous verrons que selon toutes les probabilités le globe terrestre est couvert par plus de cent mille espèces de végétaux. Chacune de ces espèces a sa patrie, son nom, sa forme, ses propriétés et ses usages. Toutes ces connaissances diverses ont leur degré d'utilité. Mais qui pourra éclaircir à nos yeux cette étude immense ? Quel guide trouverons-nous dans ce dédale effrayant ? Comment pourrions-nous arriver au point, je ne dis pas de savoir tout ce qui est relatif à l'histoire individuelle de chaque végétal, mais de pouvoir à volonté trouver ce que les autres hommes en ont su, et être par conséquent à même de savoir si ce que nous observons l'a déjà été par quelqu'autre ? Ce service éminent nous ne pouvons l'attendre que d'une méthode telle, qu'après avoir divisé successivement en plusieurs groupes ces nombreux individus du règne végétal nous arrivions par une marche sûre à connaître celui qui nous intéresse : c'est cette partie de l'étude des végétaux que je désigne sous le nom de Taxonomie botanique* ».

Nous voyons ici que le mot taxonomie désigne au départ une méthode permettant de retrouver facilement des individus après les avoir classés. Faire de la taxonomie,

c'est donc mettre de l'ordre dans la diversité du vivant. La taxonomie met en place des outils nécessaires aux botanistes.

4.1.2 Taxonomie ou taxinomie ?

On trouve, en français, deux graphies pour désigner la même discipline : *taxonomie* et *taxinomie*. Fischer et Rey (1983) expliquent que seule la forme *taxinomie* est étymologiquement correcte, et que la forme *taxonomie* vient de ce que Candolle a fait une erreur lors de sa formation en 1813. « *Le seul fait que taxonomie ait été choisi par les systématiciens et zoologistes de langue anglaise qui font autorité dans ces sciences, ne justifie pas l'erreur de Candolle ni les fausses interprétations étymologiques et définitions qu'en ont fourni ces auteurs pour cautionner l'emploi de ce terme et le préférer à taxinomie : seul ce dernier mot est correct. Aussi ce n'est pas le purisme, mais la rigueur qui conseille de parler de taxinomie* » (p. 113). Les auteurs concluent en disant qu'« *il ne nous appartient évidemment pas de décider de la forme qui sera retenue. Cela est du ressort des spécialistes de la science des classifications* ».

Plutôt que d'entrer dans le débat sans fin qui fait s'opposer les partisans d'une étymologie juste aux partisans de l'usage¹, je considérerai pour ma part que le dernier mot revient à son auteur, aussi mauvais helléniste soit-il. Je choisis donc d'utiliser le terme « taxonomie » pour la suite de ce travail.

4.2 Définitions d'auteurs

Je me suis intéressée en second lieu aux différentes définitions des termes *taxinomie*, *taxonomie* ou *taxonomy* que j'ai pu rencontrer dans la littérature scientifique. Ces définitions sont regroupées annexe D². Quelques-unes d'entre elles sont reproduites ci-après. La taxonomie est tour à tour vue comme une discipline cherchant à :

réaliser un inventaire : *Taxonomy attempts to isolate from the actual experience a sort of a photographic image of the living world, and to*

1. sur les 5 784 références de documents francophones ayant pour sujet *classification* dans le catalogue du Système Universitaire de Documentation (SUDOC, 9 millions de références d'ouvrages présents dans les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche), 55 d'entre elles contiennent le terme *taxonomie* dans leur titre, contre 19 pour le terme *taxinomie*.

2. C'est volontairement que j'ai pris en compte l'ensemble des différentes définitions rencontrées, indépendamment de leur valeur. Le but de cette démarche n'était pas de trier a priori ces définitions, en fonction d'un ou de plusieurs critères. Ce but était d'avoir un aperçu complet des différentes représentations qu'ont les scientifiques de la taxonomie.

4.2 Définitions d'auteurs

give a comprehensible description of the regularities observed in this image (Dobzhansky, 1936, p. 353);

classer : *Taxonomy, the scientific study and practice of classification (Wilson (2003, p. 78), Crisci (2006b, p. 217));*

donner des noms : *Taxonomy provides a vocabulary to discuss the world : specifically, the names of organisms (Knapp et al., 2002);*

refléter les relations évolutives : *One aim of taxonomy is to produce a system of order which will reflect as nearly as possible the evolutionary relationships between the entities being classified (Stebbins, 1969, p. 26);*

comprendre la flore et la faune : *[understanding] the immense floras and faunas of the world (Anonymous, 1946b);*

créer un système prédictif : *[taxonomy is the] science of discovering, describing and naming individual species of fauna and flora of the world, and aims to explain their relationships in order to provide a predictive classification system (Klopper et al., 2002);*

explorer la planète : *taxonomy can justly be called the pioneering exploration of life on a little known planet (Wilson, 2004);*

rendre accessible l'information : *discover and describe Earth's species; classify species according to their phylogenetic relationships; and make what is known about them accessible through Linnaean classifications and names (Wheeler, 2007; de Carvalho et al., 2008);*

théoriser la classification : *Taxonomy is the theoretical study of classification, including its bases, principles, procedures, and rules (Simpson, 1961, p. 11).*

Ces citations ne seraient pas complètes sans mentionner le terme de *systématique*. En effet, quelle que soit l'époque, on peut retrouver sous ce terme de *systématique* certaines des définitions précédentes :

systematic science - in the discrimination of species and varieties, in the study of the relations of these groups to each other and to their environment (Jordan, 1888);

Systematics, the study of biological diversity (Wilson, 1985b).

Si la synonymie entre les termes *taxonomy* et *systematics* est implicite dans les définitions précédentes, il arrive que certains auteurs la mentionnent de manière explicite :

I was therefore somewhat surprised recently to find out that I was attaching to the word in my own thoughts certain meanings which it did not hold for some other taxonomists to whom I talked. So I desire to make it clear that I am now using the terms taxonomy, classification, and systematics as absolutely synonymous, since they all inherently refer to the science of arrangement or classification (Blackwelder, 1940) ;

taxonomy and systematics, the naming and classification of living thing (Thiele et Yeates, 2002).

En revanche, d'autres auteurs font une distinction nette entre les deux :

Systematic Zoology, as I perceive it, consists of three main fields, taxonomy, nomenclature and phylogeny. These fields are constantly playing upon each other and being played upon by other fields of zoology and by other sciences (Metcalf, 1937, p. 514) ;

Taxonomy provides the bricks, and systematics the plan, with which the house of the biological sciences is built (May, 2004).

Nous voyons que de multiples définitions plus ou moins restrictives de ce domaine de recherche ont toujours existé, et qu'elles cohabitent encore aujourd'hui. Essayons maintenant de les analyser plus finement.

On peut envisager qu'un domaine de recherche puisse être défini en faisant appel à cinq caractéristiques : ses objets d'étude, ses buts, ses paradigmes, ses méthodes ou encore ses productions. En s'intéressant aux définitions précédentes, on se rend compte que les auteurs mobilisent tour à tour quatre d'entre elles : ses objets d'étude, ses buts, ses paradigmes et ses productions ; et que chacune de ces caractéristiques varie en fonction de l'auteur qui l'emploie.

Lorsque les objets d'étude sont précisés, ils sont tour à tour des organismes [7, 11, 14, 18, 20, 21, 23, 25, 30, 31, 40, 43, 44, 49, 66, 71, 72]³, des êtres vivants [2, 68, 24, 26, 46, 55], leurs relations [12, 16, 20, 27, 34, 45] ou encore, depuis les années 2000, la biodiversité [36, 47, 50, 53, 42, 52] ; terme qui semble avoir remplacé les mots plus anciens de *floras and faunas* [6], *forms* [5], *kinds* [65].

Curieusement, certaines de ces définitions proposent les résultats du travail taxonomique comme objet d'étude : ce peuvent être des espèces [3, 8, 15, 17, 27, 28, 37, 38, 48, 54, 33], la classification [9], les taxons [10, 19, 22] ou leurs concepts [12].

La majorité des définitions (2/3) fait appel à l'un de ces mots : classer [1, 4, 9, 14, 15, 17, 18, 20, 23, 24, 26, 30, 31, 41, 44, 45, 46, 49, 64, 66, 68, 70, 71, 72], décrire

3. Ces numéros renvoient aux citations listées en annexe D.

4.2 Définitions d'auteurs

[2, 3, 6, 17, 22, 23, 27, 29, 35, 38, 39, 40, 43, 44, 46, 49, 53, 72, 55], nommer [6, 8, 22, 25, 26, 27, 30, 31, 38, 40, 43, 46, 49, 50, 53, 54, 66, 68, 71, 55] ou identifier [8, 22, 28, 38, 37, 39, 41, 43].

Si on ajoute celles qui contiennent les mêmes notions sous des formes différentes [5, 10, 11, 13, 21, 33, 34, 36, 42, 48, 52], on trouve que 87 % des définitions proposées se fondent sur les productions de la taxonomie, c'est-à-dire que quasiment 9 définitions sur 10 en donnent une vision utilitaire. Cela est indépendant de l'époque considérée. Les définitions restantes sont très générales, basées sur les termes de « *compréhension* » [7, 12], « *traitement* » [16], « *exploration* » [32], « *recherche* » [47].

Les paradigmes sont très peu utilisés dans ces définitions. Il s'agit de classer en fonction des « *relations naturelles présumées* » [14], des « *relations phylogénétiques* » [38] ou des « *parentés évolutives* » [10, 70].

Suite à cette analyse on dégage deux observations importantes :

1. il n'existe pas de définition consensuelle de la taxonomie ; elles vont de la plus large (« *biodiversity research* » [47]) à la plus restrictive (« *expertise in the identification of [...] species* » [37]) ;
2. deux types de définitions cohabitent. Le premier type présente la taxonomie comme un prestataire de services (réaliser un inventaire, classer, donner des noms, créer un système prédictif, rendre accessible l'information) tandis que le second la présente comme une science à part entière, avec ses propres questionnements et paradigmes (refléter les relations évolutives, comprendre la flore et la faune, explorer la planète, théoriser la classification) ;
3. néanmoins, quelle que soit l'époque considérée, la taxonomie est majoritairement perçue comme une discipline produisant les données nécessaires aux autres sciences et la société.

Étant donné que le mot *classification* revient dans 23 définitions, soit 41 % des définitions répertoriées, on pourrait penser qu'il est pertinent de définir la taxonomie comme le travail de classification des organismes. Or ce mot, *classer*, est un mot polysémique : il peut être attribué à trois activités différentes. En fonction de l'emploi qu'on en fera, on aura donc, là encore, des sens différents pour le terme *taxonomie*. Il peut en effet s'agir de :

- trouver la place d'un objet inconnu de nous dans une hiérarchie préexistante (synonyme : identifier) ;
- créer la place d'un objet inconnu dans une classification préexistante (synonyme : décrire et nommer une nouvelle espèce) ;
- réorganiser les différentes partitions d'une classification préexistante entre elles, sans modifier les partitions initiales (synonyme : créer de nouveaux taxons).

On peut résumer cela à « Assigner des individus à des partitions », i.e. valider un classement préexistant ; « Créer des partitions », ou « Hiérarchiser ces partitions »,

i.e. remettre en question un classement préexistant. Les outils mis en oeuvre ne sont pas les mêmes suivant que l'on travaille dans l'un ou l'autre cas : la description d'espèces nouvelles fait appel à des compétences et des paradigmes différents de la création d'une hiérarchie des espèces.

Dans ce cadre, différents termes ont également émergé pour traduire cette distinction : d'un côté microtaxinomie (Mayr et Ashlock, 1991), taxinomie α ou eidonomie (Dubois, 2008a; Dubois et Raffaëlli, 2009) pour désigner les descriptions, et de l'autre macrotaxinomie (Mayr et Ashlock, 1991) pour désigner la création d'une hiérarchie⁴.

4.3 Définition par la pratique

Haas et Häuser ont cherché à définir ce qu'est un taxonomiste. Ils ont proposé neuf définitions différentes (tableau 4.1), ainsi que les problèmes soulevés par chacune d'elle. Aucune n'est complètement satisfaisante : « *If you have further definitions, I would be happy to include them!* » .

SOMEONE WHO	LIMITATIONS
describes species	but birds and mammals are almost all described
revises taxa	but taxa with well established taxonomy are rarely revised
works Code compliant	but it is difficult to establish if a work is Code compliant
has an academic education in taxonomy	but amateurs, but lack of university courses
has comprehensive knowledge of a taxon	but ecologists, ethologists and physiologists have comprehensive knowledge and will not consider themselves as taxonomist
can identify a group of organisms	but has no formal education and little knowledge of the Code
is paid for taxonomic work	but amateurs and university scientists, who are employed for teaching and non-taxonomic skills
is member of a systematic society	but supporters join these societies without having extensive knowledge and not all taxonomists join
is listed in a directory	but not all like to be registered

TABLEAU 4.1 – Proposition de définitions d'un taxonomiste par Haas et Häuser (2006).

4. Les termes cladonomie (Dubois, 2008a; Dubois et Raffaëlli, 2009; Brummitt, 1997; Dubois, 1997) ou phylonomie (Dubois, 2008a; Dubois et Raffaëlli, 2009) que l'on peut également rencontrer dans la littérature ne peuvent être retenus comme synonymes de macrotaxinomie car ils sont en réalité restreints à un type particulier de macrotaxinomie : la macrotaxinomie cladistique.

4.4 Absence de consensus

Nous voyons donc que plusieurs définitions de la taxonomie cohabitent, et que le choix de l'une ou l'autre définition (restreinte ou large, par exemple) aura des répercussions manifestes sur d'éventuelles mesures de la discipline. Doit-on considérer que la classification des espèces, ou l'étude des phylogénies, est partie prenante de la taxonomie, ou bien qu'elle est un domaine en émergence, qui s'en différencie peu à peu ? Doit-on considérer que les travaux portant sur la distribution des espèces, sans données supplémentaires sur leur identité, sont de la taxonomie, ou bien que ce sont de l'écologie ?

Ces questionnements nous renvoient à la notion de spécialisation des sciences. L'histoire naturelle étant l'une des disciplines scientifiques les plus anciennes, c'est elle qui a donné naissance aux autres disciplines biologiques. Partant de là, il peut être difficile de délimiter son propos, que ce soit par la raison ou par consensus. Nous reconnaissons ici la problématique classique de la taxonomie, mais appliquée aux champs de recherche scientifiques : comment classer des entités suite à un processus d'évolution ? De manière peu étonnante, on constate que la métaphore de la taxonomie comme une race ou une espèce (le plus souvent pour dire qu'elle est en voie d'extinction) est ancienne : « *the race of naturalists became nearly extinct* » (Merriam, 1893) et régulièrement reprise (Shear, 1934; Isely, 1972; Kemp, 1989; Mayr, 1989; Woodland, 2007; Bateman, 2008).

Etant donné que les frontières de la taxonomie sont difficiles à poser, j'ai trouvé intéressant de considérer l'ensemble de la zoologie comme référentiel disciplinaire pour la suite de ce travail et de chercher à établir ses frontières en étudiant les liens existant entre les travaux de l'ensemble de la zoologie. J'ai considéré que la zoologie comprenait toute discipline dont les études portent sur des animaux, i.e. toute discipline ayant pour sujet d'étude un organisme nommé suivant l'*International Code of Zoological Nomenclature*. Cette approche me semblait d'autant plus pertinente qu'elle permettait d'apporter des données supplémentaires en faveur de cette définition de la discipline, plutôt que d'embrasser un parti-pris a priori.

La partie suivante va donc tenter d'établir ces frontières en s'intéressant aux relations entretenues par la taxonomie avec les autres disciplines zoologiques. J'ai cherché à déterminer la place concrètement prise par les différents éléments mentionnés précédemment – description des espèces, données de distribution, phylogénies, ... – au sein de l'ensemble des productions écrites de zoologie, en m'attardant sur deux questions : (1) ces différents éléments forment-ils au sein de la zoologie un domaine clairement défini, que l'on peut appeler « taxonomie » ? Si oui, (2) quelles sont les interactions que ce domaine entretient avec les autres disciplines ?

Ces questions peuvent trouver une réponse au travers de la cartographie des sciences,

champ de recherche en plein essor au carrefour entre bibliométrie, visualisation de l'information, étude des systèmes complexes ou encore statistique.

4.5 Vers une cartographie de la zoologie

Comme le dit Velt dans son article récapitulatif *Cartographies des connaissances scientifiques* (2011), le classement des savoirs est une question philosophique majeure dont on peut trouver des représentations graphiques très anciennes. Citons ainsi, pour le plaisir des yeux, les représentations synthétiques présentes dans l'*Arbor Scientiae* de Lulle (1515, figure 4.1), dans les *Tableaux accomplis de tous les arts libéraux* de Savigny (1587, figure 4.2), ou encore dans l'introduction du premier tome de l'*Encyclopédie* de Diderot et d'Alembert (1751, figure 4.3). La plupart du temps, ces connaissances sont présentées sous forme d'arbre, i.e. sous une forme hiérarchisée. Ce choix de représentation laisse la porte ouverte au progrès puisqu'il rend possible l'ajout de nouvelles disciplines. La représentation sous forme de cercle de Savigny (figure 4.2) est une exception, qui ne laisse pas de place au progrès (Bibliothèque nationale de France, 1996). La forme hiérarchique prévaudra jusqu'au milieu du XX^e siècle, et sera choisie en 1876 par le bibliothécaire américain Melvil Dewey pour mettre en place un système de rangement des ouvrages encore présent dans les bibliothèques du monde entier : la classification décimale de Dewey.

Si elles facilitent l'accès aux connaissances et aux documents, ces représentations hiérarchiques utilitaires ne permettent toutefois pas de rendre compte des relations que les disciplines ont entre elles. Au milieu du XX^e siècle, des représentations tenant compte de ces relations voient alors le jour : celle du physicien John Desmond Bernal en 1939 (Bernal, 1939) et celle du chimiste Harold Ellingham en 1948 (figure 4.4). Ces cartes sont basées sur les connaissances personnelles que chacun de leurs auteurs possèdent vis-à-vis de leur discipline.

Dans les années 1960 apparurent les premières cartes basées non plus sur une expertise, mais sur des données quantitatives collectées, analysées puis représentées visuellement (Velt, 2011). Les données servant de support à la construction de ces cartes sont d'abord les données de co-citation, répertoriées dans les grandes bases de données bibliographiques disponibles (voir page 6). Le principe est que les documents ou journaux souvent cités ensemble, i.e. souvent co-cités, définissent un ensemble de recherche cohérent (Small, 1973) qui peut être visualisé graphiquement. Par la suite d'autres types de données furent utilisés pour représenter les liens entre disciplines : données linguistiques (Callon *et al.*, 1986), association des deux méthodes précédentes

4.5 Vers une cartographie de la zoologie

ou, encore plus récemment, traces informatiques laissées par les chercheurs lors de leurs recherches bibliographiques sur des sites dédiés (Bollen *et al.*, 2009).

Depuis les années 2000, l'accroissement des moyens de calcul et le développement concomitant des outils a fait de la représentation cartographique des sciences une discipline en pleine expansion, d'autant plus investie que ses problèmes rejoignent ceux d'initiatives plus larges telles que la cartographie du *World Wide Web* ou des réseaux sociaux virtuels. La cartographie scientométrique ne se développe plus seulement dans les laboratoires, mais également au sein d'entreprises, véritables *start-up* montées par des chercheurs ; au sein d'associations ou encore de consortiums. Nous pouvons ainsi citer l'exposition *Places and spaces*⁵, qui donna lieu à la publication de l'*Atlas of science* (Börner, 2010), les *Maps of science*⁶ de SciTech Strategies (Boyack, Klavans et Small), ou encore les cartes de la science européenne développées par *WebAtlas* (Franck Ghitalla). Le tableau serait incomplet s'il ne mentionnait les projets de cartographies des sciences basés non plus sur la description de leur structure, mais sur la compréhension de leur(s) dynamique(s) et de leur évolution, comme le projet de cartographie des controverses scientifiques *Macospol* (coordinateur : Bruno Latour) ; ou encore plus récemment, les projets étudiant la cartographie comme outil d'évaluation de la recherche (Rafols *et al.*, 2010).

Nous voyons que ce domaine présente un éventail d'outils très divers. Pour répondre aux deux questions précédemment posées, j'ai sélectionné parmi toutes ces techniques deux approches cartographiques différentes et complémentaires.

J'ai choisi d'utiliser le *Zoological Record* comme source principale des données. Cette base ne contenant pas de données de citations, j'ai utilisé l'approche relationnelle par cooccurrences de mots-clés pour réaliser une carte de la zoologie et ainsi répondre à la question de la délimitation de la taxonomie.

Pour répondre à la seconde question, portant sur les liens entre les disciplines, j'ai choisi cette fois de faire appel aux données de citation présentes dans une base moins exhaustive, le *Web of Science*. Je me suis donc intéressée aux liens de citation existant au sein d'un ensemble de documents portant sur un taxon étudié par plusieurs disciplines. Je suis partie de l'idée que chaque auteur, par l'acte de citation,

5. « Cette initiative scientifique vise à inspirer un débat pluridisciplinaire sur le meilleur moyen de suivre et de transmettre l'activité humaine et le progrès scientifique à l'échelle mondiale. Elle est composée de deux éléments : une exposition des cartes pour les conférences et les manifestations éducatives, avec des impressions grands formats, et l'équivalent numérique sur <http://scimaps.org>, contenant une sélection de cartes et d'explications détaillées sur leur origine et leurs auteurs [...] L'initiative se déroule sur 10 ans, avec une dizaine de nouvelles cartes par année, pour un total de 100 cartes en 2014 » (http://scimaps.org/static/docs/P&S_info_french_web_2011.pdf).

6. « SciTech Strategies, Inc. publishes disciplinary, competency, and paradigm maps. Its disciplinary maps are used to provide a basic understanding of how science is organized; competency maps are used to identify the interdisciplinary strengths of a research, lab, institution, region, or nation; and paradigm maps are used for planning research efforts in a specific area of research. The company was founded in 1991 and is based in Berwyn, Pennsylvania » (<http://investing.businessweek.com/research/stocks/private/snapshot.asp?privcapId=109115333>).

situe son article dans un contexte de travail et un environnement thématique particulier, faisant ainsi jonction (ou rupture) avec les thématiques pré-existantes dans lesquelles il évolue. En mettant à jour les liens de citation existant entre différents articles, je cherche donc à visualiser l'espace thématique au sein desquels ces auteurs placent leur travail, ainsi que les liens qui s'établissent, par là même, entre ces différents environnements. Je ne fais ici aucune hypothèse sur la « valeur scientifique » ou qualitative des articles ainsi cités.

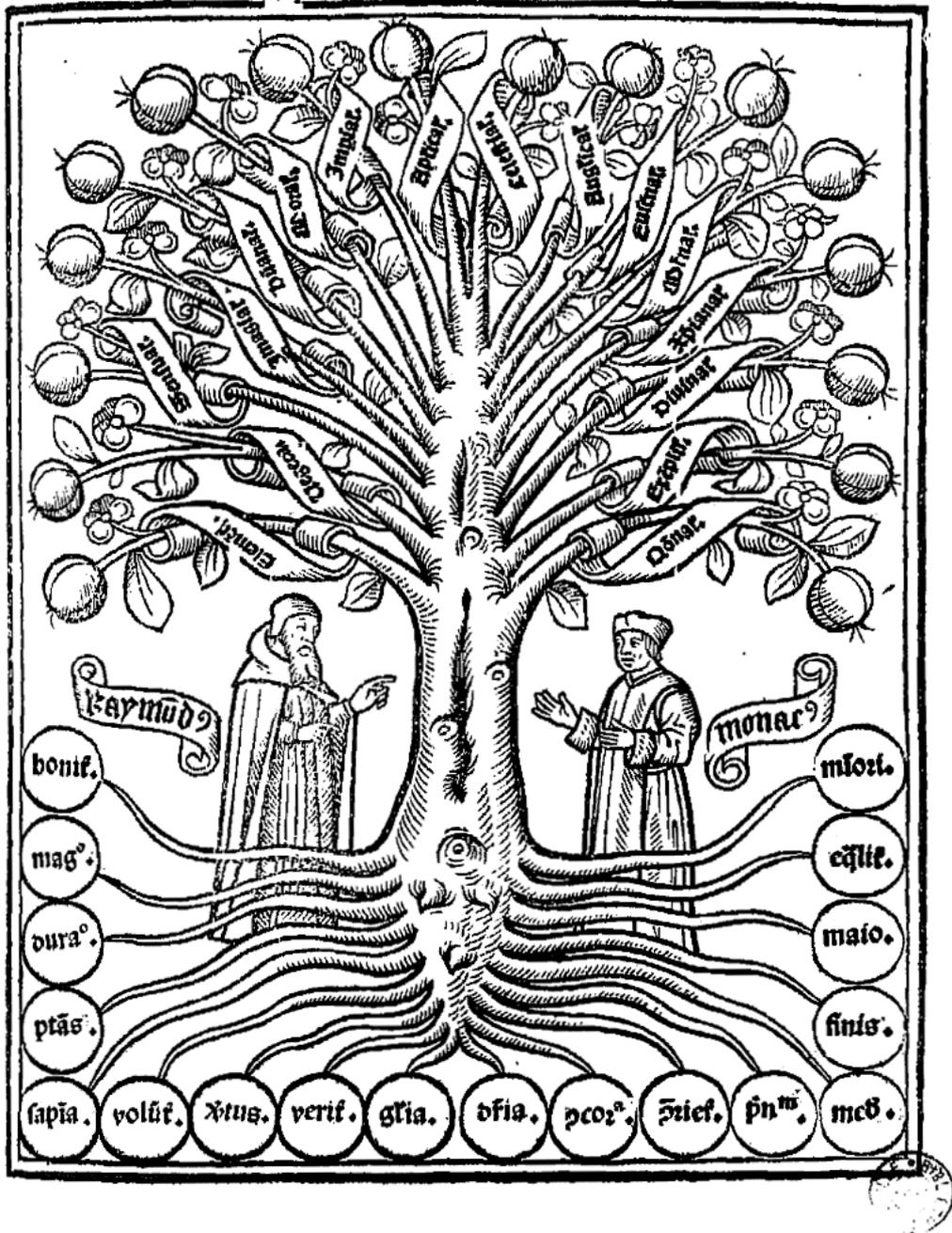


FIGURE 4.1 – L'Arbor scientie de Lulle (1515). Source : Gallica. « Dans l'Arbre de sciencia, composé à Rome en 1295 [...], les dix-huit racines de l'arbre des sciences sont constituées par les neuf principes transcendants (ou dignités divines) et par les neuf principes relatifs de l'Art (differentia, concordantia, contrarietas ; principium, medium, finis ; maioritas, aequalitas, minoritas). L'arbre se subdivise en seize branches, donc chacune correspond à l'un des arbres qui formeront la forêt de la science : l'arbor elementalis, l'arbor vegetalis, (botanique et application de la botanique à la médecine), sensualis (êtres sensibles, sentants et animaux), imaginalis (ceux parmi les êtres mentaux qui sont semblables aux êtres réels considérés dans les arbres précédents), humanalis, moralis (éthique, doctrine des vices et des vertus), imperialis (lié à l'arbor moralis [...]), apostolicus (gouvernement ecclésiastique et hiérarchie de l'Église), celestialis (astronomie et astrologie), angelicus (les anges et les serviteurs angéliques), eternalis (immortalité, monde ultra-terrestre, enfer et paradis), maternalis (mariologie), christianus (christologie), divinus (théologie, dignités divines, substance et personnes de Dieu, perfections et productions divines) » (Rossi, 1993, p. 60).

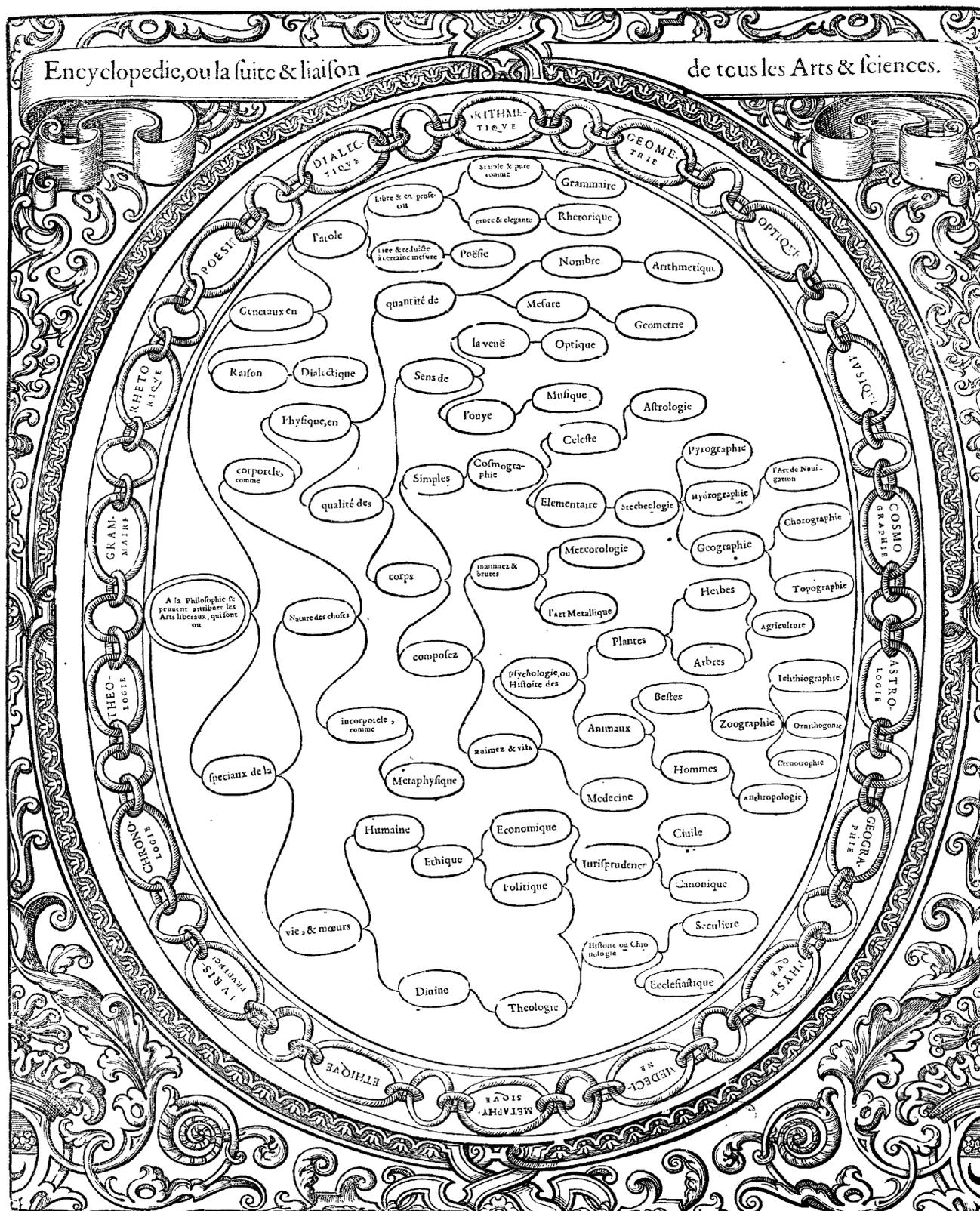


FIGURE 4.2 – Encyclopedie tirée des Tableaux accomplis de tous les arts libéraux de Savigny (1587). Source : Gallica. On distingue : Grammaire, Rhétorique, Poésie, Dialectique, Arithmétique, Géométrie, Optique, Musique, Cosmographie, Astrologie, Géographie, Physique, Médecine, Métaphysique, Éthique, Jurisprudence, Chronologie, Théologie. « Les animaux sont ou irraisonnables & bestes, ou raisonnables & hommes : La description des bestes appartient généralement à la Zoographie, à laquelle se referent l'Icthyographie ou description des poissons, l'Ornithogonie ou maniere de nourrir les oiseaux, Ctenotrophie ou pastorale » (Savigny, 1587).

* SYSTÈME FIGURÉ DES CONNOISSANCES HUMAINES.

ENTENDEMENT.

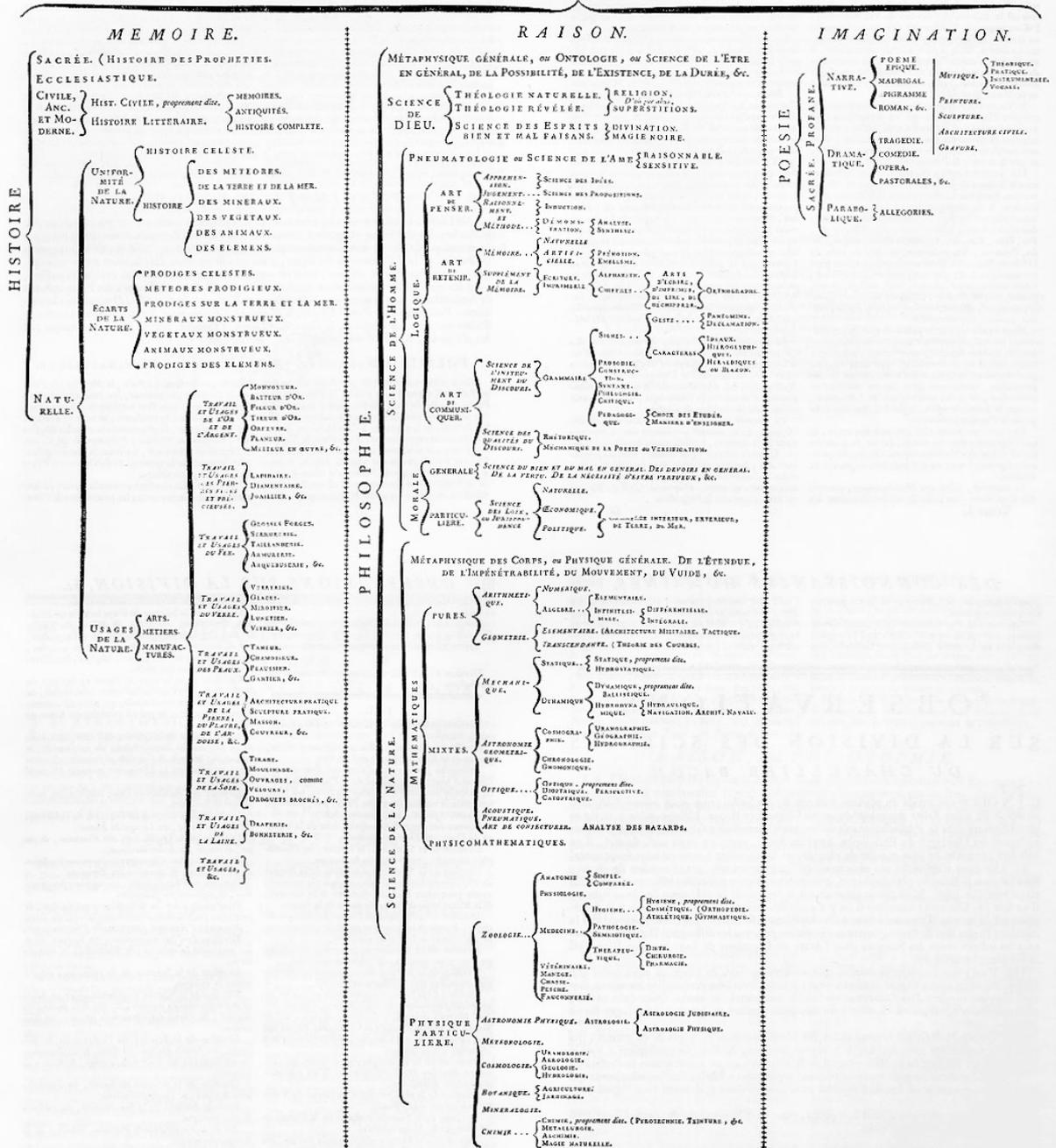


FIGURE 4.3 – Système figuré des connaissances humaines tiré du Discours préliminaire des éditeurs de Diderot et d'Alembert (1751). Source : ARTFL Encyclopédie Project.

A CHART ILLUSTRATING SOME OF
THE RELATIONS BETWEEN THE BRANCHES OF NATURAL SCIENCE AND TECHNOLOGY
H.J.T. Ellingham. 1948.

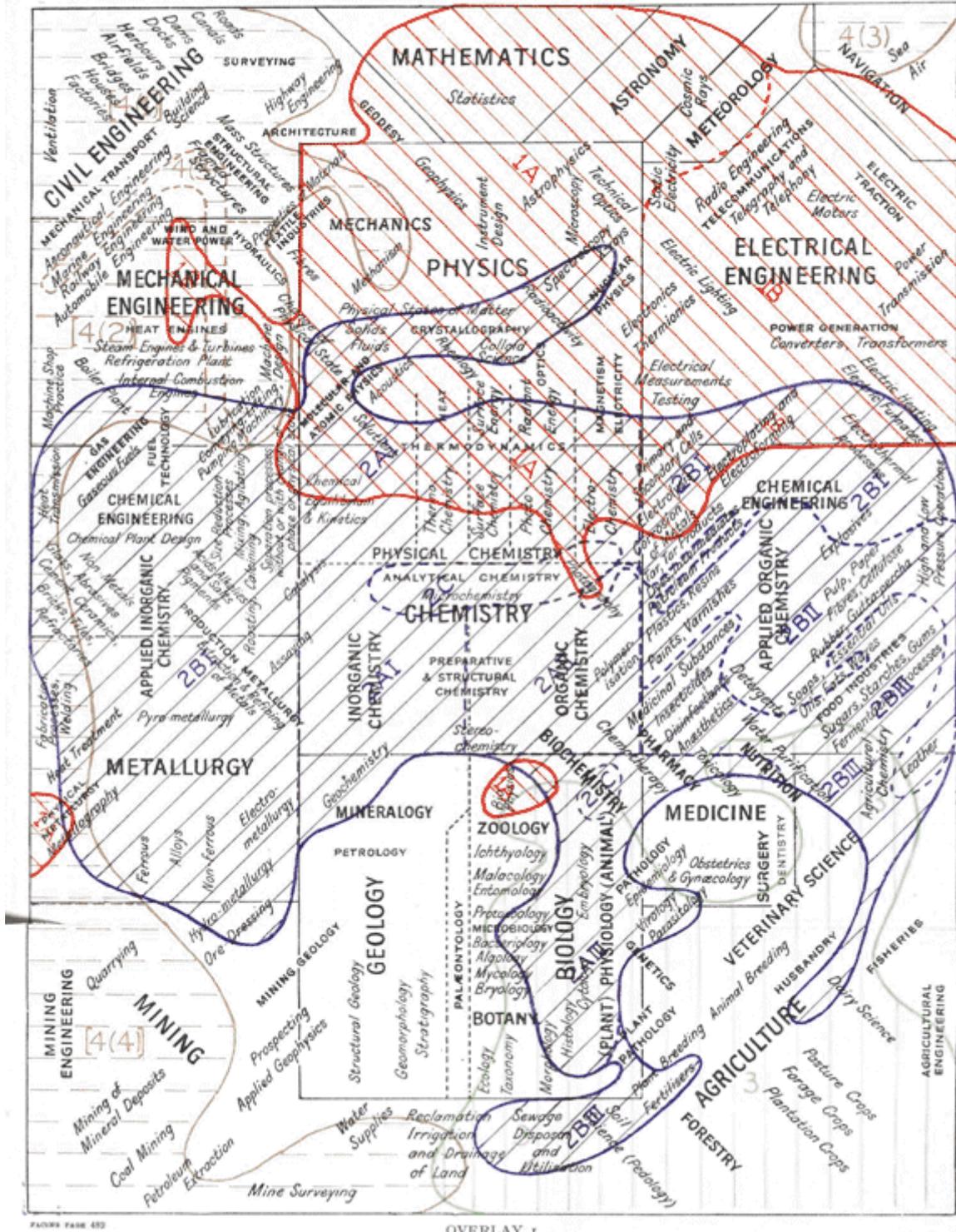


FIGURE 4.4 – Carte des sciences naturelles et technologiques d'Ellingham (1948). Avec l'aimable autorisation de la Royal Society.

Chapitre 5

Cartographies de la zoologie

5.1 Méthodologie

5.1.1 Principe général

Le processus général de cartographie d'un domaine scientifique à partir de liens de cooccurrence a été résumé par Börner *et al.* (2003) :

data extraction : extraction des données à partir des grandes bases de données bibliographiques ;

unit of analysis : choix de l'unité d'analyse : journal, document, auteur, terme... Les données sont vérifiées pour en ôter les doublons et erreurs typographiques diverses ;

measures : calcul de la fréquence d'apparition de chaque terme et de chaque paire de termes ; que ceux-ci soient des auteurs, des mots-clés, des citations, etc. ;

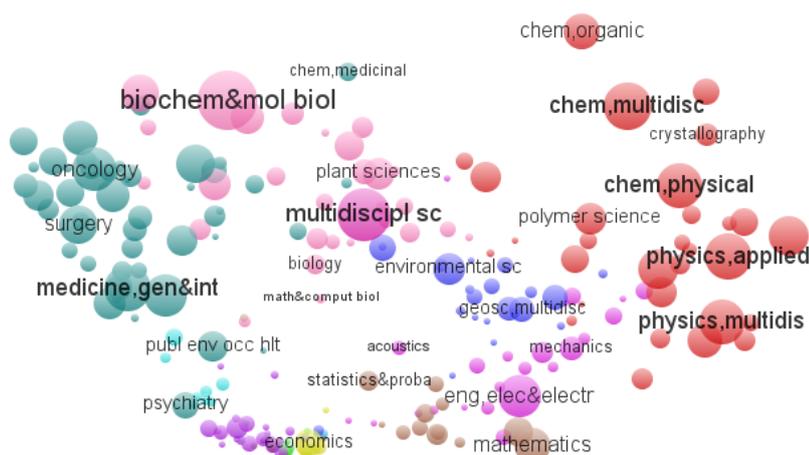
layout : normalisation des mesures d'occurrence et de cooccurrence. Il existe différentes méthodes de normalisation des occurrences : *td-idf*, *latent semantic analysis*... et différentes méthodes de normalisation des cooccurrences, ou mesures de similarité : *association strength*, *cosine*, *inclusion index*, *Jaccard index*... Elles ne seront pas détaillées ici, on pourra se référer à Colliander et Ahlgren (2011) pour plus d'informations.

Actuellement, deux types de cartes sont communément réalisées en recherche bibliométrique (van Eck et Waltman, 2010) :

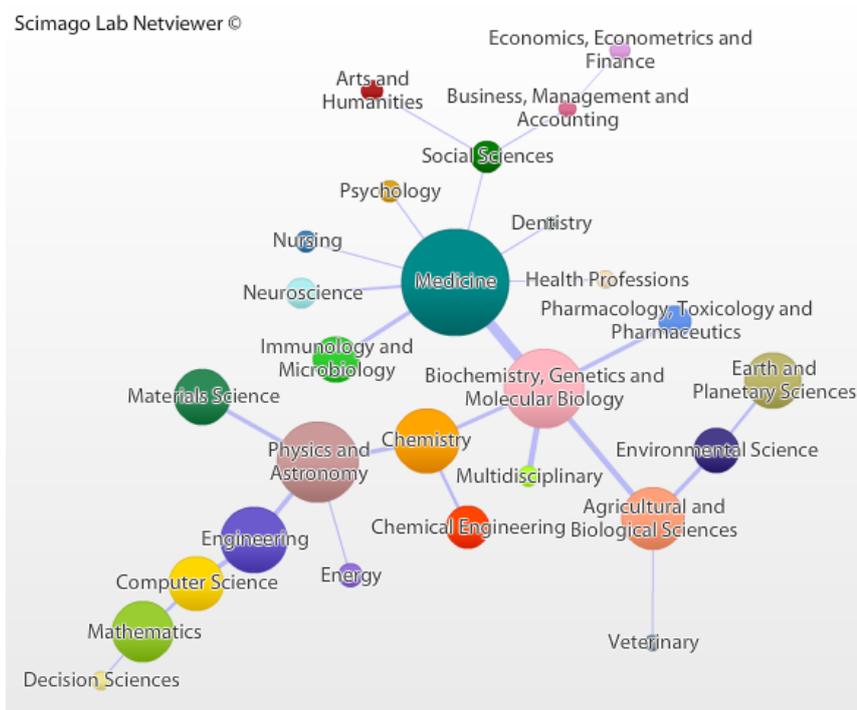
les cartes de distance (*distance-based maps*) : la distance entre deux nœuds reflète la force de la relation qui existe entre eux. Plus la distance est courte, plus la relation est forte (figure 5.1a). Par exemple, deux auteurs peuvent être proches sur ces cartes car ils collaborent toujours ensemble.

Elles mettent en jeu des méthodes telles que : *multidimensional scaling*, *principal component analysis*, *visualization of similarities*, etc. ;

les cartes de graphes (*graph-based maps*) : dans ce cas, la distance n'est plus liée à la force de la relation et celle-ci est modélisée par une ligne entre les deux individus (figure 5.1b). Ces cartes permettent d'éviter que les nœuds se chevauchent, mais elle rend les regroupements plus difficiles à détecter. Elles mettent en jeu des algorithmes de positionnement tels que : *Kamada-Kawai*, *Fruchterman-Reingold*, *Force Atlas*, *OpenOrd*, etc.



(a) distance-based map. Illustration tirée de van Eck et Waltman (2009).



(b) graph-based map. Illustration construite à partir de Scimago (2007) (données de citation pour la France, 2007-2008).

FIGURE 5.1 – Exemples des deux principaux types de cartes existant en scientométrie.

5.1 Méthodologie

Ces cartes se retrouvent également sous les termes de *cartes de nature statistique* pour les premières, et *cartes de nature connective* pour les secondes (Moya Anegón *et al.*, 2006).

display : exploration de la carte avec des outils informatiques dédiés : recherche, zoom, application de filtres, etc.

5.1.2 Outils disponibles

Si les premières cartographies étaient réalisées à l'aide d'outils tels que Pajek (sorti en 1996), des outils de cartographie « prêts à l'emploi » sont désormais disponibles. Une revue d'un certain nombre de ces outils de cartographie des sciences a été réalisée récemment par Cobo *et al.* (2011). Ils sont présentés succinctement tableau 5.1. Ces différents outils sont loin d'être équivalents puisqu'ils ont pour la plupart été développés dans des optiques différentes. Ainsi, CiteSpace (Chen, 2006) se concentre sur la détection des domaines scientifiques en émergence, VOSviewer (van Eck et Waltman, 2010) est dédié à la représentation graphique et à la navigation dans des cartes de grande taille ; SoNIA (Bender-deMoll et McFarland, 2006) à la visualisation dynamique des réseaux, i.e. à leur évolution au cours du temps ; tandis que l'Overlay Toolkit (Rafols *et al.*, 2010) permet de comparer la place prise par différentes disciplines au sein de plusieurs corpus issus du *Web of Science*.

Software tool	Last version	Date	Developed by
Pajek	2.05	2011-25-09	Ljubljana univ (Slovenia)
Bibexcel	2011-28-09	2011-28-09	Umeå univ (Sweden)
CiteSpace	3.0 R2	2011-09-28	Drexel univ (USA)
Leydesdorff's Software	N/A	N/A	Amsterdam univ (The Netherlands)
Network Workbench Tool	1.0.0	2009-09-15	Indiana univ (USA)
Sci ² Tool	0.5.1	2011-05-05	Indiana univ (USA)
VOSviewer	1.4.0	2011-09-10	Leiden univ (The Netherlands)
Visone	2.6.3	2010-09-13	Konstanz univ (Germany)
SoNIA	1.1	2004-09-18	Stanford univ (USA)
Overlay Toolkit	N/A	N/A	NSF project <i>Measuring and Mapping IDR</i>
Gephi	0.8 alpha	2011-04-08	Gephi Consortium (France)

TABLEAU 5.1 – *Logiciels d'étude de réseaux scientométriques en accès libre. Modifié à partir de Cobo et al. (2011).*

L'accès aux outils disponibles est l'une des plus grandes difficultés rencontrées lors d'un travail scientométrique. Qu'il s'agisse de chaînes logicielles montées à partir d'outils existants (Access, Endnote, Sphinx...), ou d'outils spécifiques, développés dans des buts précis par des équipes d'ingénieurs ou de chercheurs indépendants, il semble n'y avoir que peu de discussion ou de tentative de concertation entre développeurs. La plupart des articles du domaine ne détaillent pas l'aspect logiciel de leurs analyses. À titre d'exemple, c'est en 2010 que paraît le premier article (Mallig,

2010) proposant de manière détaillée une marche à suivre pour construire une base de données à des fins bibliométriques. Alors même que la bibliométrie existe depuis 60 ans, et les bases de données relationnelles depuis... 50 ans.

5.1.3 Réalisation des cartes

Étant donné le nombre de références que ce travail pouvait potentiellement mobiliser (plus de 76 000 références pour la seule année 2008), j'ai choisi de réaliser trois cartes différentes de la zoologie, sur des taxons particuliers : Aves, Amphibia et Hemiptera. Le premier (oiseaux) correspond à un taxon considéré comme très bien connu d'un point de vue taxonomique ; le deuxième (amphibiens) a été choisi comme exemple de taxon vertébré et le troisième (punaises, etc.) comme exemple de taxon d'invertébrés. Les deux derniers taxons ont un volume de publications similaire et ont donc été choisis pour cette raison. J'ai privilégié les cartes de distance et la méthodologie défendue par *van Eck et al. (2010)* : *association strength* pour la normalisation des données et *visualisation of similarities* pour la spatialisation. Ces deux approches ont été implémentées dans le logiciel VOSviewer.

Les trois corpus de références ont été extraits du *ZR* grâce aux requêtes suivantes :

ST = Aves AND PY = 2008
 ST = Amphibia AND PY = 2008
 ST = Hemiptera AND PY = 2008

Pour chaque corpus, les données ont été importées dans **R** et les mots-clés disciplinaires ont été extraits du champ DE (Descriptors Data) grâce à un script dédié. Les mots faisant référence à un taxon ou à une région ont été mis de côté. J'ai ensuite utilisé le logiciel Calliope pour vérifier la liste de mots-clés et réindexer les documents. Un certain nombre de mots ont été mis en synonymie. Par exemple, ont été considérés comme synonymes :

acoustic signals
bioacoustic data
bioacoustic variation
 ...

ou encore

mitochondrial genome
MTDNA
mitochondrial gene sequences

5.2 Résultats

mitochondrial markers *MTDNA analysis*

...

Lors de la réindexation, le seuil d'occurrence a été choisi à 1 ‰ de l'ensemble des références, soit 8 pour les données sur les oiseaux et 3 pour les données sur les hémiptères et amphibiens. Cela a porté le nombre total de mots à 1 600 environ pour chacune des trois cartes. Le seuil de cooccurrence a été choisi à 1. Le réseau des cooccurrences a ensuite été construit avec Sci², la matrice de similarité calculée avec VOSviewer et la visualisation réalisée avec VOSviewer (options par défaut).

Les résultats sont présentés figures 5.2 à 5.7. Plus deux mots sont proches sur la carte et plus leur similarité est élevée ; c'est-à-dire qu'on les retrouve fréquemment associés. Chaque point de la carte possède une couleur qui dépend de la densité des mots qui l'environnent et du poids de ces mots, i.e. de l'ensemble des liens qu'il possède. Plus le nombre de mots environnants est élevé et leur éloignement faible, plus la couleur du point tend vers le rouge. Cette couleur tend également vers le rouge si ces mots ont un poids élevé. Pour résumer, plus la couleur se rapproche du rouge et plus les domaines concernés sont riches en vocabulaire et fortement connectés. La taille des mots est proportionnelle à leur poids, i.e. l'ensemble des liens qu'ils possèdent.

5.2 Résultats

La carte 5.2 issue des références portant sur Aves est structurée en trois grands domaines distincts : écologie, études acoustiques, études de génétique moléculaire. Cette structuration se retrouve sur la carte 5.4 concernant les amphibiens. Le profil est toutefois différent en ce qui concerne la carte 5.6 des hémiptères : si l'écologie domine largement, les deux autres domaines qui s'en détachent sont la description de nouvelles espèces et la répartition des taxons.

Ces trois cartes nous montrent que la description des nouvelles espèces et la dénomination des nouveaux taxons forme une discipline en marge des autres sous-champs de la zoologie. Les phylogénies et données de distribution sont situées dans le voisinage de cette discipline mais s'en distinguent. Les données de distribution (« new_record », « checklist »...) font le lien avec l'écologie tandis que les données de phylogénie, très liées à l'utilisation de techniques moléculaires, font le lien avec les études portant sur l'évolution et les processus adaptatifs. Je n'ai pas séparé les articles de paléontologie des articles portant sur les faunes actuelles. Leurs travaux n'étant pas réalisés sur la base de caractères moléculaires, on pourrait penser qu'ils exercent une influence sur la topologie de la carte, en accentuant l'éloignement entre

les travaux de phylogénie moléculaire et ceux de descriptions d'espèce. Cependant, leur nombre étant faible (1 % pour Aves, Hemiptera et 2 % pour Amphibia), j'ai considéré que cette influence pouvait être négligée.

Nous voyons que ces cartes dessinent la taxonomie comme la discipline traitant de la description de nouvelles espèces et attribuant des noms aux taxons supra-spécifiques. Du point de vue des publications, les descriptions de nouvelles espèces et dénominations de nouveaux taxons forment donc une discipline à part. Cela définit la taxonomie comme la discipline centrée sur la nomenclature. Or, nous savons que ce qui se situe au cœur de la nomenclature est l'emploi de spécimens types : les noms d'espèce ne sont pas définis par une liste de caractéristiques données (morphologiques, ...) mais ils sont rattachés à un spécimen particulier, appelé type. Ces spécimens au rôle important sont conservés dans les collections d'histoire naturelle. On retrouve donc, à travers ces cartes, la place primordiale qu'occupent les collections en taxonomie. Une nouvelle définition de la taxonomie peut être donnée, qui soit volontairement dégagée de toute allusion à un paradigme, une technique ou des productions attendues : ceux-ci ont varié au cours du temps, et sont appelés à varier de nouveau. Je propose de la définir en fonction de ses objets d'étude, de manière plus précise que ce qui a été proposé dans les définitions précédentes, comme suit : **discipline étudiant de manière comparée des spécimens de collections naturalistes à des fins nomenclaturales**. Ainsi, **tout travail faisant référence à des spécimens de collections naturalistes étudiés à des fins nomenclaturales est un travail taxonomique**.

On pourrait objecter que cet isolement est artificiel : si phylogénies et descriptions sont déconnectées, c'est que les règles de publication elle-mêmes, et en particulier l'évaluation par l'impact factor, pousse les taxonomistes à scinder leurs résultats en deux, publiant les phylogénies dans les revues à impact factor élevé et les descriptions dans des revues à faible impact factor. Cet argument est fréquemment retrouvé dans la littérature (voir par exemple [Werner, 2006](#)). Si cette hypothèse est juste, on ne devrait pas trouver de taxonomistes spécialisés uniquement dans l'une ou l'autre de ces pratiques.

J'ai testé cette idée en dénombrant les auteurs ayant uniquement décrit des espèces nouvelles, uniquement réalisé des phylogénies ou ayant réalisé les deux, sur quatre ans de publications traitant d'Amphibiens dans le *Zoological Record*. J'ai extrait du *Zoological Record* les références des travaux portant sur les amphibiens et publiés de 2005 à 2008 ($n = 9\ 071$), par la requête

$$ST = \text{Amphibia AND PY} = 2005-2008$$

J'ai ensuite importé les données dans **R**, et fait un travail important d'homogénéisation des auteurs (voir page 33). J'ai ainsi mis en évidence que environ 13 %

5.2 Résultats

des auteurs étaient des synonymes. J'ai enfin réalisé différents dénombrements. Les résultats sont présentés tableau 5.2.

Nombre total d'auteurs	12 398
Nombre d'auteurs décrivant de nouvelles espèces	542
Nombre d'auteurs publiant des phylogénies	1 054
Nombre d'auteurs publiant nouvelles espèces et phylogénies	168

TABLEAU 5.2 – *Dénombrement d'auteurs travaillant sur les amphibiens en 2005-2008, par spécialité. Source : Zoological Record.*

Le tableau 5.2 montre que 31 % des auteurs de nouvelles espèces participent à la publication de phylogénies, tandis que 16 % des auteurs de phylogénies participent à la publication de nouvelles espèces. Par conséquent, plus de deux tiers des auteurs de nouvelles espèces ne font pas de phylogénies, et plus de quatre phylogénéticiens sur cinq ne décrivent pas de nouvelles espèces. On trouve donc des scientifiques spécialisés dans l'une ou l'autre de ces pratiques. La définition donnée précédemment à la discipline reste donc valable.

Voyons à présent si cette délimitation, observée au niveau des publications et des acteurs, se retrouve également au niveau des citations. Les publications portant sur des collections se citent-elles uniquement entre elles? Sont-elles citées par d'autres types de publications?

5.2 Résultats

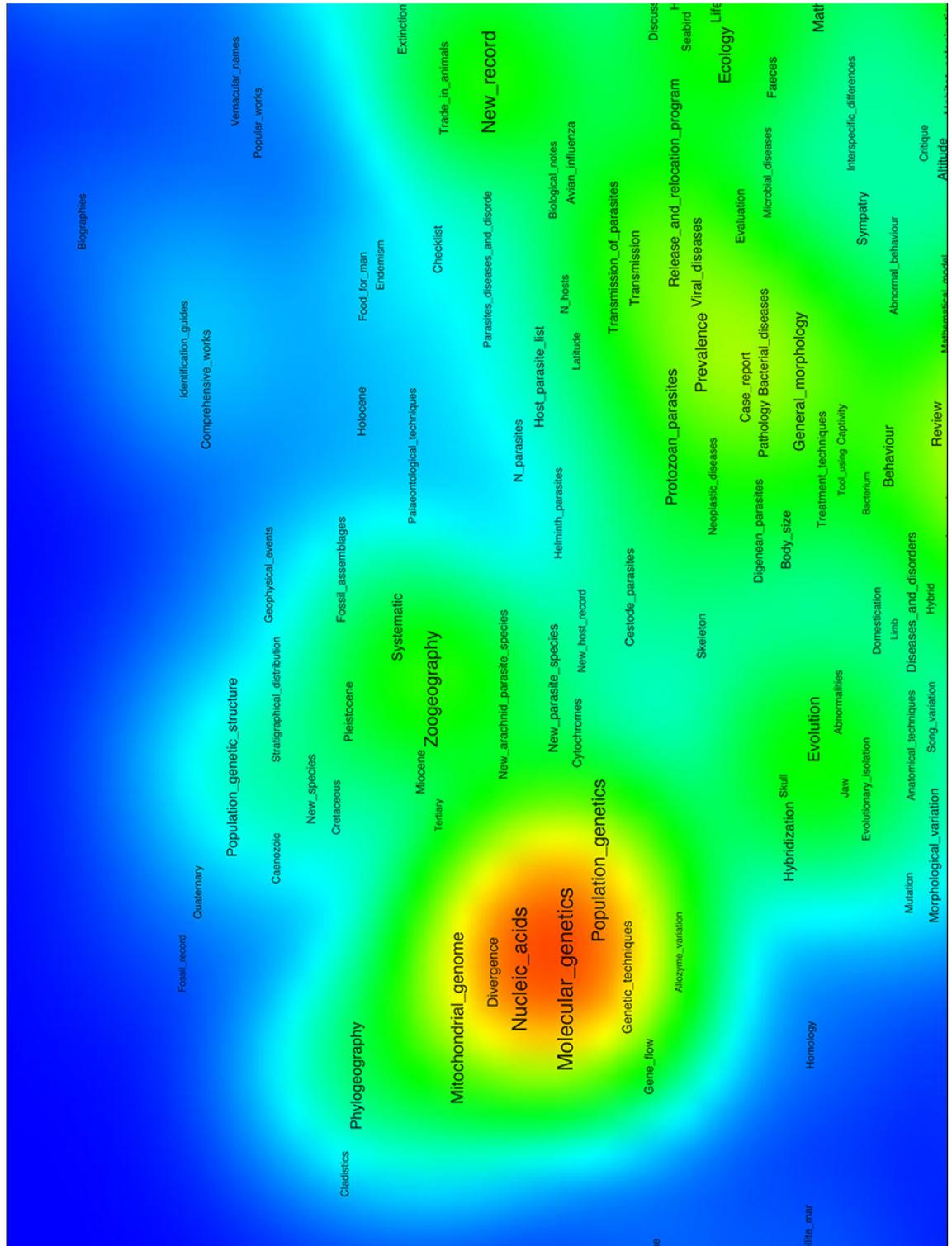


FIGURE 5.3 – Représentation cartographique des études zoologiques concernant le taxon *Aves* en 2008 – Détail : la taxonomie. Champ DE du Zoological Record, visualisation avec VOS-viewer.

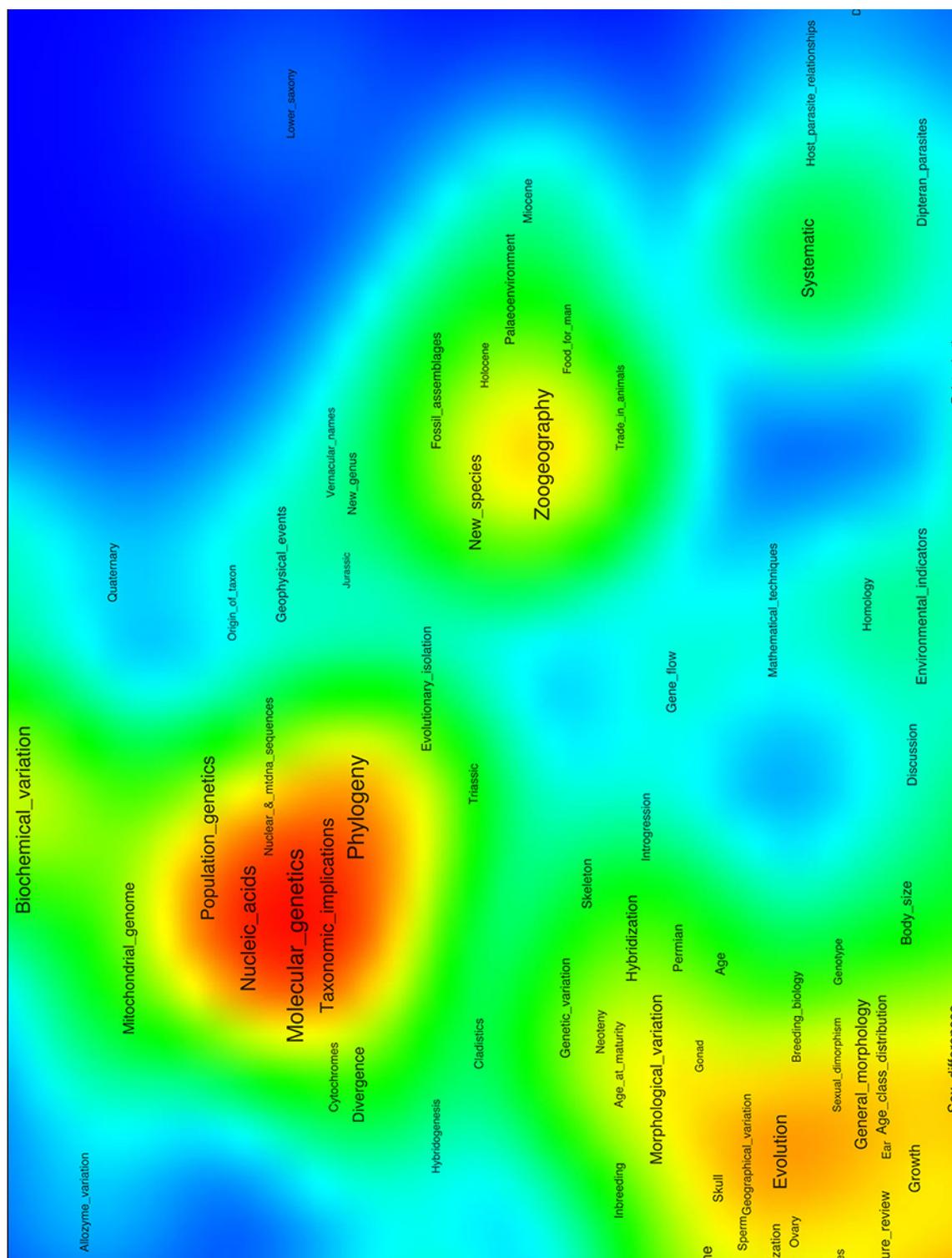


FIGURE 5.5 – Représentation cartographique des études zoologiques concernant le taxon *Amphibia* en 2008 – Détail : la taxonomie. Champ DE du Zoological Record, visualisation avec VOSviewer.

Chapitre 6

Cartographie de citations au sein de la zoologie

6.1 Méthodologie

6.1.1 Choix du corpus et du *WoS*

Je voulais pouvoir définir de façon claire pour chaque article s'il portait ou non sur des collections. J'ai donc choisi de travailler sur un nombre restreint d'articles. Le choix a été fait de travailler sur un taxon précis : les Pelobatidae, une famille d'Amphibiens qui est étudiée dans d'autres disciplines que la systématique.

Le *ZR* ne possède pas de données de citation entre les articles de sa base. Il fallait donc avoir accès à une autre base de données pour accéder à ce type d'information. Il existe seulement deux bases de données bibliographiques qui proposent des données de citation : le *WoS* de Thomson Reuters, et depuis 2004, *Scopus* d'Elsevier. Le choix de travailler sur le *WoS* a été déterminé par des contraintes logistiques ; aussi je n'entrerai pas dans la comparaison de leurs avantages et inconvénients respectifs. Pour plus d'information, on pourra consulter [Bosman *et al.* \(2006\)](#) et [Falagas *et al.* \(2008\)](#).

La première contrainte logistique est liée à l'accès à ces bases. Celui-ci étant conditionné à la souscription d'un abonnement coûteux, la plupart des établissements de recherche sont amenés à choisir entre l'une et l'autre de ces bases. Le muséum et le CNRS n'ont jusqu'à présent pas fait exception à la règle et ils ont fait le choix de proposer à leur personnel de recherche un accès au *WoS*. Ainsi j'ai eu uniquement accès au *WoS*. Il faut noter que les modalités d'abonnement diffèrent entre les deux structures : celui du MNHN couvre la période de 1956 jusqu'au présent, tandis que celui du CNRS court de 1900 à aujourd'hui.

La seconde contrainte est liée à la disponibilité des outils d'analyse. Le *WoS* étant la base la plus ancienne, elle dispose d'un plus grand nombre d'outils d'analyse dédiés que Scopus, et notamment d'un outil d'analyse de graphes de citation, HistCite.

6.1.2 Estimation de la couverture taxonomique du *WoS*

J'ai cherché à estimer le nombre de revues de taxonomie présentes dans le *Web of Science*.

Méthodologie

J'ai cherché toutes les références de travaux comportant des descriptions de nouvelles espèces en utilisant dans le *ZR* la requête :

$$SD = ("New species" OR "Sp nov") AND PY = 2008$$

J'ai également cherché toutes les références de travaux comportant des phylogénies en utilisant dans le *ZR* la requête :

$$SD = ("Phylogeny" OR "Host parasite phylogeny") AND PY = 2008$$

Dans les deux cas, j'ai filtré les résultats de manière à ne retenir que les références d'articles et j'ai utilisé l'outil *Analyze results by Source Titles* pour obtenir la liste des revues.

La liste des revues présentes dans le Journal Citation Reports est disponible en ligne à l'adresse http://admin-apps.webofknowledge.com/JCR/JCR?RQ=TITLES_FULL.

Les données ont ensuite été importées et homogénéisées sous **R**.

Résultats

NOMBRE TOTAL DE :	
revues dans le <i>Web of Science</i>	6 620
articles décrivant au moins une nouvelle espèce (<i>ZR</i>)	5 661
revues concernées (<i>ZR</i>)	804
articles publiant des phylogénies (<i>ZR</i>)	3 002
revues concernées (<i>ZR</i>)	708

TABLEAU 6.1 – Estimation de la couverture taxonomique du *Web of Science* en 2008. Sources : *Web of Science*, *Zoological Record* (*ZR*).

6.1 Méthodologie

En 2008, 26 % des revues publiant des descriptions de nouvelles espèces et 46 % des revues publiant des phylogénies sont dans le *Web of Science*. Ces chiffres s'accroissent jusqu'à 34 % et 52 % en 2010 (sachant qu'ils sont probablement sous-estimés étant donné que l'indexation des documents de 2010 dans le *ZR* est incomplète).

Ces chiffres sont à manier avec précaution. En effet, étant donné que la distribution des articles dans les revues suit une loi de Bradford (voir [Distributions bibliométriques : Lotka, Bradford, Zipf page 9](#)), on peut s'attendre à ce qu'ils varient fortement si l'on pouvait compter non pas le nombre de revues, mais le nombre d'articles présents dans le *Web of Science*. Si les 65 % de revues non indexées sont situées dans la zone de dispersion, cela signifierait que la taxonomie est tout de même très présente dans le *Web of Science*. Le tableau 6.2 présente les dix premières revues qui contiennent le plus grand nombre d'articles décrivant de nouvelles espèces en 2008. Parmi elles, la moitié seulement étaient indexées cette année-là.

NOM	NOMBRE D'ARTICLES	INDEXÉ
ZOOTAXA	830	OUI
ACTA ZOOTAXONOMICA SINICA	124	NON
JOURNAL OF PALEONTOLOGY	58	OUI
JOURNAL OF NATURAL HISTORY	56	OUI
JOURNAL OF PARASITOLOGY	55	OUI
PROC. OF THE ENTOM. SOC. OF WASHINGTON	48	OUI
ANNALES ZOOLOGICI WARSAW	46	NON
PALAEONTOLOGY OXFORD	46	NON
PALEONTOLOGICHESKII ZHURNAL	44	NON
GENUS WROCLAW	43	NON

TABLEAU 6.2 – Liste des 10 premières revues publiant des descriptions d'espèces en 2008. Source : Zoological Record. Nombre d'articles : nombre d'articles décrivant de nouvelles espèces.

Essayons d'estimer ce nombre d'articles. La requête

TS = ("new species" OR "sp. nov." OR "sp nov") AND PY = 2008 AND
Document Types = (Article)

Refined by : *Web of Science* Categories = (ZOOLOGY OR ENTOMOLOGY OR
MARINE FRESHWATER BIOLOGY OR PALEONTOLOGY OR
PARASITOLOGY)

estime le nombre d'articles de description présents dans le *Web of Science*. Elle renvoie aussi certains travaux qui mentionnent le mot-clé *new species* dans leur titre alors qu'il s'agit simplement d'une nouvelle espèce pour une localité, et non une description de nouvelle espèce. Cependant ces articles sont peu nombreux (2,3 % en

2008 dans le *ZR*)¹. Cette requête nous donne ainsi une première idée de la disparité précédemment mentionnée. Elle nous renvoie 2 427 articles. Il y aurait environ 43 % des articles de description de nouvelles espèces de 2008 présents dans le *Web of Science*. Ce nombre passerait à 52 % en 2010.

Malgré le fait que ces chiffres soient à prendre avec précaution, on peut noter qu'ils ne sont pas négligeables et qu'il n'est par conséquent pas contre-indiqué de se servir du *Web of Science* pour obtenir des données sur la taxonomie. Notons que je n'ai pas fait la distinction entre les nouvelles espèces fossiles et les nouvelles espèces actuelles. Dans le *WoS*, la paléontologie représente 12 % des articles du corpus précédent ; que ce soit en 2008 ou en 2010.

6.1.3 Définition des mots-clés de recherche

J'ai détaillé dans la section [Construction d'un corpus de travail issu du *ZR*](#) (page 29) la construction d'une requête sur cette thématique, pour le *ZR*. J'ai repris les mêmes mots-clés pour le *WoS*, en supprimant toutefois le mot *amphibia* puisqu'il n'y a pas d'indexation nomenclaturale dans le *WoS* :

TS = ("pelobatidae" or "pelobates" or "megophryidae" or "pelodytidae" or "scaphiopodidae" or "borneophrys" or "brachytarsophrys" or "leptobranchella" or "leptobranchium" or "leptolalax" or "megophrys" or "ophryophryne" or "oreolalax" or "scutigera" or "xenophrys" or "pelobates" or "pelodytes" or "scaphiopus" or "spea")

On obtient 991 références. Dans le cas du *WoS*, on trouve que *spea* peut être lié à *multicriteria optimization, the Strength Pareto EA (SPEA)* ou *allotype speA*. Les résultats sont donc filtrés en excluant les disciplines suivantes :

Refined by : [excluding] Subject Areas = (MICROBIOLOGY OR IMMUNOLOGY OR INFECTIOUS DISEASES OR ENGINEERING, ELECTRICAL & ELECTRONIC OR COMPUTER SCIENCE, THEORY & METHODS OR COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE).

On obtient alors 721 références au 26 avril 2010. Parmi celles-ci, 118 faisaient seulement indirectement référence aux Pelobatidae (il s'agissait, par exemple, d'études sur les Salamandres pour étudier une problématique déjà étudiée sur les Pelobatidae) et 88 faisaient encore référence aux sujets homonymes précédemment mentionnés. Je n'ai donc conservé que 515 articles dont la distribution temporelle est présentée figure 6.1.

1. Rapport du résultat de la requête « SD = ("New species" OR "Sp nov") AND PY = 2008 » sur celui de la requête « TS = ("New species" OR "Sp nov") AND PY = 2008 ».

6.1 Méthodologie

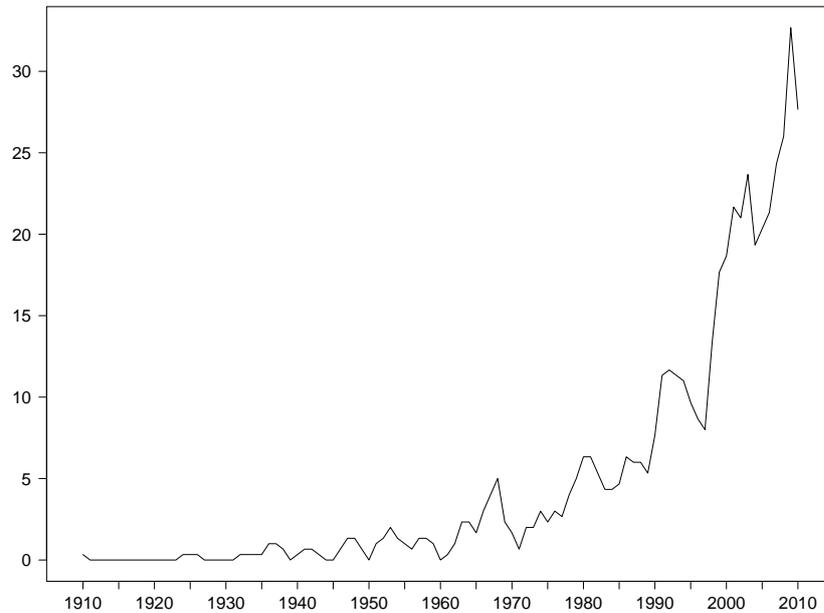


FIGURE 6.1 – *Distribution des travaux traitant de Pelobatidae et enregistrés dans le Web of Science. Données du 26 avril 2010. Moyenne mobile sur 3 ans.*

6.1.4 Construction du graphe

Il existe un outil un peu particulier, lancé officiellement par Eugene Garfield en 2007 et aujourd'hui distribué par *Thomson Reuters* : *HistCite*. Il permet de réaliser des historiographes, c'est-à-dire de visualiser de manière chronologique les liens de citations entre documents portant sur un même sujet (voir figure 6.2).

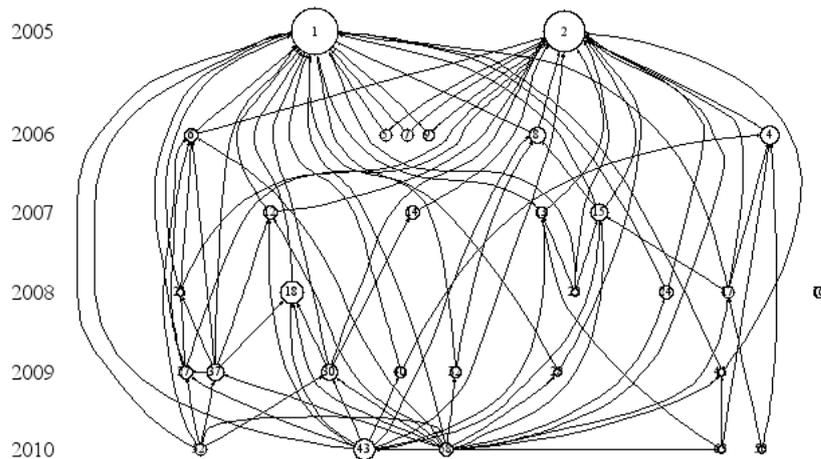


FIGURE 6.2 – *Exemple d'historiogramme réalisé avec HistCite : historiogramme des 30 articles les plus cités sur la notion de « taxonomie intégrative », fondée en 2005 par Dayrat. Chaque nœud représente un article. La taille des nœuds est proportionnelle au nombre de citations reçues par l'article au sein des 105 documents. Données : Web of Science. Requête TS = "integrative taxonomy" réalisée le 06.10.2011 (n = 105 références).*

Les données du *Web of Science* sont exportées sous format balisé, dont on peut trouver un exemple figure 6.3. Comme pour le *ZR*, l'export est limité à 500 références par fichier. J'ai ensuite importé les données dans HistCite et réalisé l'historiographe, dont le pas à pas est détaillé en annexe E. Le graphe est ensuite visualisé dans Gephi à l'aide d'un algorithme de spatialisation. J'ai choisi ForceAtlas (Jacomy *et al.*, 2011) pour sa clarté de résultat. ForceAtlas est un algorithme de dessin basé sur les forces (*force-based* ou *force-directed algorithm*). Il simule un système de forces entre les noeuds qui apparente les liens à des ressorts et les noeuds à des particules de même charge. Les noeuds sont déplacés suivant les règles de physique classique jusqu'à atteindre un état stable. Concrètement, chaque noeud représente un article portant totalement ou partiellement sur la famille des Pelobatidae. Chaque lien correspond à une citation d'un article envers l'autre. Le réseau est dirigé puisqu'il rend compte du sens de la citation. Les agrégats rendus visibles par Gephi sont interprétés comme des environnements thématiques qu'une analyse ultérieure, basée sur les métadonnées des articles (titre, résumé, adresses, etc.) vise à identifier, décrire et nommer.

Le graphe peut se lire à plusieurs niveaux : au niveau des taxons étudiés, de la thématique de questionnement, du pays dans lequel se déroule l'étude ; et également d'un point de vue temporel. La lecture est excessivement complexe et sera présentée ici de manière synthétique.

6.2 Résultats

Le graphe (figure 6.4) présente quatre zones principales, numérotées ici de 1 à 4. Les clusters 1 et 4 correspondent à des thématiques portant sur l'évolution et l'écologie ; le cluster 2 à des études de physiologie et le cluster 3 à des études de taxonomie. On distingue, en rouge, les travaux portant sur des collections. Les données de collections concernent les études sur le développement squelettique et la classification. On remarque trois groupements différents : le premier (en haut) porte sur la classification de parasites de Pelobatidae. Le second porte sur la classification d'un sous-groupe de Pelobatidae et le troisième (en bas) sur des études de développement squelettique. Les articles de phylogénie, ici représentés en bleu, font le lien entre taxonomie alpha et les autres études sur l'évolution. Les données de distribution trouvent leur place dans la thématique écologique et, en outre, ne contiennent pas de données de collection.

On voit donc clairement, ici, que la taxonomie α est située dans des clusters de citation isolés. Ce graphe thématique renforce la définition donnée en préambule. Il existerait une discipline clairement définie, travaillant *sur des collections* à des fins *nomenclaturales*.

6.2 Résultats

L'isolement thématique important qui existe entre la taxonomie et les autres disciplines (alors que les autres disciplines ont de nombreux liens entre elles) peut s'expliquer par le fait que ce ne sont pas les mêmes taxons qui sont étudiés. Il s'agit en effet principalement de *Leptolalax* pour la taxonomie et de *Scaphiopus* et *Spea* pour les autres disciplines. Nous verrons dans le chapitre [Mesure de l'apport de la taxonomie aux autres disciplines](#) (page 126) que des taxons étudiés taxonomiquement ne seront pas étudiés simultanément dans d'autres disciplines.

Maintenant que nous avons circonscrit la taxonomie, tentons de mesurer son déclin.

```

FN ISI Export Format
VR 1.0
PT J
AU Cai, B
    Bu, WJ
AF Cai, Bo
    Bu, Wenjun
TI A review of Yemmatropis (Hemiptera: Lygaeoidea: Berytidae), with
    descriptions of two new species from China
SO ZOOTAXA
LA English
DT Article
DE Hemiptera; Berytidae; Yemmatropis; review; new species; China
AB Four species of the genus Yemmatropis Hsiao, 1977, from China are
    treated in this paper, in which Yemmatropis erectus Henry, 2002 is
    newly recorded from China. Yemmatropis hsiao sp. nov. and Yemmatropis
    nigrovittatus sp. nov. are described. Photographs of head and pronotum
    and drawings of openings of male genital capsules and parameres of all
    four species, and a key to all Yemmatropis species, are provided to
    assist in identification. The type specimens of the new species are
    deposited in the Institute of Entomology, College of Life Sciences,
    Nankai University, Tianjin, China.
C1 [Cai, Bo; Bu, Wenjun] Nankai Univ, Coll Life Sci, Inst Entomol,
    Tianjin 30071, Peoples R China.
RP Bu, WJ, Nankai Univ, Coll Life Sci, Inst Entomol, 94 Weijin Road,
    Tianjin 30071, Peoples R China.
EM wenjunbu@nankai.edu.cn
FU Natural Science Foundation of China [30725005, 31071959, J0930005]
FX We express our thanks to Ms. Huiqin Yue (Nankai University, China) for
    her technical assistance in taking photographs. We thank Dr. Peter
    Kment (National Museum, Czech Republic) for his helpful comments on our
    manuscript. This project was supported by Natural Science Foundation of
    China (No. 30725005, No. 31071959, and No. J0930005).
CR ASHLOCK PD, 1957, ANN ENTOMOLOGICAL SO, V50, P407
    HENRY TJ, 1997, CONTRIBUTIONS AM ENT, V30, P1
    HENRY TJ, 1997, MEMOIRS ENTOMOLOGICA, V19, P1
    HENRY TJ, 1998, CONTRIBUTIONS AM ENT, V30, P1
    HENRY TJ, 2002, SPECIES DIVERSITY, V7, P165
    HSIAO TY, 1974, ACTA ENTOMOL SINICA, V17, P55
    HSIAO TY, 1977, HDB DETERMINATION CH, V1, P283
    PERICART J, 2001, CATALOGUE HETEROPTER, V4, P232
    REN SZ, 1988, INSECTS MT NAMJAGBAR, P111
    STUSAK JM, 1989, ACTA ENTOMOL BOHEMOS, V86, P111
NR 10
TC 0
PU MAGNOLIA PRESS
PI AUCKLAND
PA PO BOX 41383, AUCKLAND, ST LUKES 1030, NEW ZEALAND
SN 1175-5326
J9 ZOOTAXA
JI Zootaxa
PD APR 4
PY 2011
IS 2808
BP 41
EP 48
PG 8
SC Zoology
GA 745FD
UT ISI:000289150400004
ER

EF
    
```


Troisième partie

Un déclin relatif

Chapitre 7

Que mesurer ?

7.1 La notion de déclin

Nous n'avons pas encore vu précisément ce que recouvre le terme de *déclin* de la taxonomie. Celui-ci peut prendre plusieurs formes. Il peut s'agir : d'une baisse du nombre de taxonomistes (Wiegand, 1936; Smith *et al.*, 2008), d'un désintérêt de la part des non taxonomistes, et plus particulièrement des autres disciplines (Metcalf, 1946; Knapp *et al.*, 2002), de l'acquisition d'une mauvaise réputation (Hitchcock, 1916; Wheeler *et al.*, 2004), d'une diminution de ses financements (Fosberg et Diehl, 1946; Anonymous, 2008b), d'une augmentation de l'âge moyen des systématiciens (Oliver, 1988; Anonymous, 2008b), d'un manque de personnel pour les collections (Silbermann, 1835; Smith *et al.*, 2008), d'un manque de crédit pour le fonctionnement des collections, etc. Comme nous le voyons, la liste est longue.

Dans ce contexte, nous voyons qu'il y aurait potentiellement plusieurs voies à explorer pour quantifier ce déclin : d'une part, celle de ses investissements ; d'autre part, celle de ses productions. Bien qu'il soit aisé de comprendre comment les investissements peuvent être mesurés (en termes de budget alloué aux collections, aux postes de taxonomistes, aux missions...), ce sont des mesures qu'il est quasiment impossible de réaliser à grande échelle aujourd'hui.

Lorsqu'il s'agit de mesurer les productions, les choses se compliquent. Qu'entend-t-on par productions ? S'agit-il d'articles, de nouvelles espèces ? De la collecte de nouveaux spécimens ? De la réalisation de missions ? On peut considérer que la collecte de nouveaux spécimens et les missions font partie de l'investissement mis en œuvre pour la discipline. Restent le nombre d'articles et de nouvelles espèces.

Pour tenter de mesurer le déclin en termes de production, je propose donc de mesurer le nombre de publications de taxonomie et le nombre d'espèces et sous-espèces décrites au cours du temps. Je propose également de mesurer de manière indirecte

les investissements faits dans la discipline, en dénombrant les revues publiant des travaux de taxonomie et les auteurs publiant des travaux de taxonomie. Il ne s'agit pas ici de faire une mesure des financements de la discipline – cela n'est vraiment pas possible dans ces conditions – mais plutôt une mesure des forces disponibles.

7.2 Définition du corpus

Comme pour la partie précédente, j'ai choisi d'utiliser le *Zoological Record* comme source principale des données. Un premier problème se pose alors : comment délimiter les articles de taxonomie dans l'ensemble de cette base ? Nous avons vu précédemment (section *Le thésaurus du ZR*, page 23) que le *Zoological Record* possède un thésaurus de thématiques scientifiques : le *Subject thésaurus*. Le premier niveau de ce thésaurus est constitué de 43 termes. La figure 7.1 présente la distribution des publications de 2008 pour chacun de ces termes. Les chiffres sont issus d'un compte de présence, ce qui signifie que si un document est attribué à plusieurs disciplines, il comptera une fois dans chaque discipline. Les définitions et périodes de validité des 7 premiers termes sont donnés tableau 7.1. Le détail de l'ensemble des 43 termes est donné dans l'annexe F. On observe que certaines spécialités sont beaucoup plus présentes que d'autres. Ainsi, *Ecology* arrive en tête, suivie de *Habitat*, *Behaviour* et *Systematics*.

<i>Ecology</i>	1864- <i>now</i>	Interrelationships between animals and their environment
<i>Habitat</i>	1864- <i>now</i>	Site and particular type of local environment occupied by an organism; restricted to items where the habitat itself is of significance
<i>Behaviour</i>	1864- <i>now</i>	Any observable action or response of an animal; includes general ethology
<i>Systematics</i>	1864- <i>now</i>	-
<i>Abiotic factors</i>	1864- <i>now</i>	External/environmental/administered influencing factors present in the wild or captive environment of an organism
<i>Nutrition</i>	1864- <i>now</i>	All aspects of nourishment, includes any mechanisms involved in the procurement, ingestion and digestion of food to provide energy for maintenance of life
<i>Evolution</i>	1864- <i>now</i>	Cumulative change in the characteristics of organisms or populations during successive generations related by descent

TABLEAU 7.1 – *Description et période de validité des 7 principaux termes de premier niveau du subject thésaurus du ZR en 2008.*

Nous voyons tableau 7.1 que la spécialité qui nous intéresse ici, *Systematics*, n'est pas définie. J'ai placé en annexe G la hiérarchie des termes qu'elle contient. Il s'agit

7.2 Définition du corpus

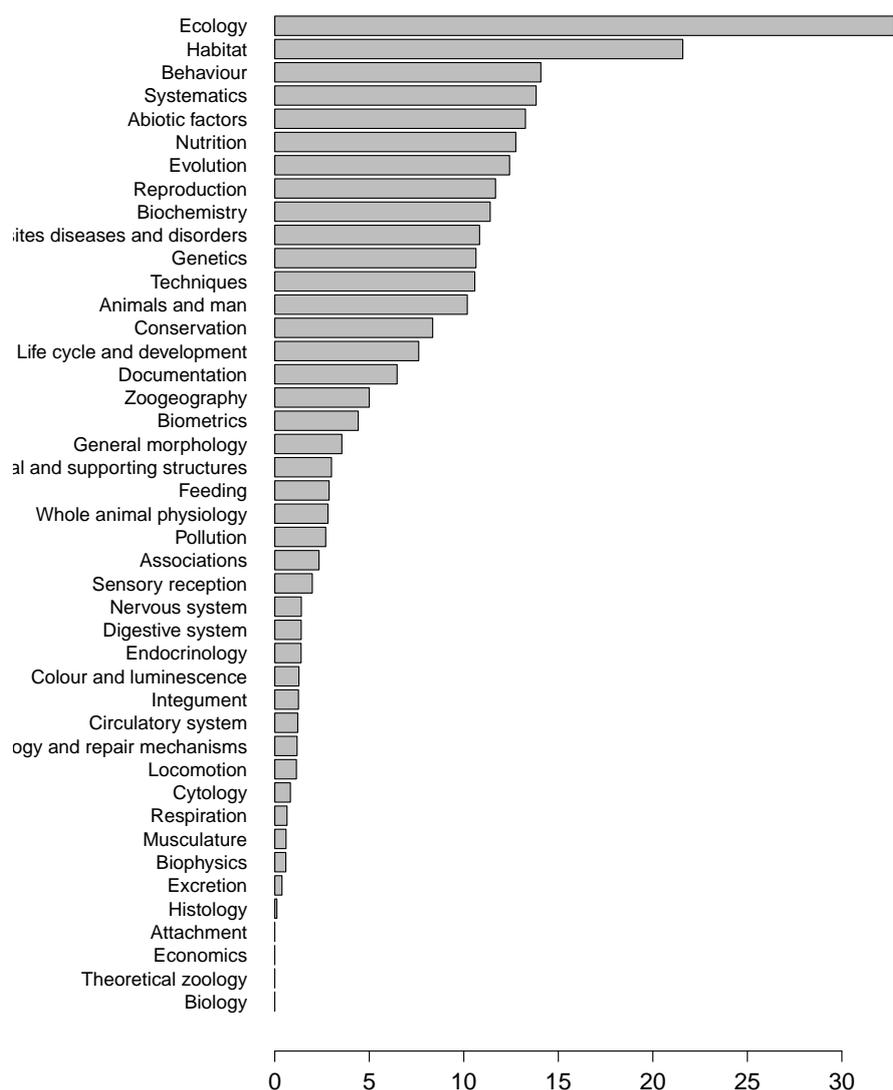


FIGURE 7.1 – Répartition des documents publiés en 2008 et présents dans le Zoological Record suivant les 43 spécialités définies par le premier niveau de son Subject thésaurus. Pourcentages de comptes de présence.

essentiellement de termes faisant référence à des actes de nomenclature : *Phylum nov*, *Junior synonym*, etc. Cela signifie que la catégorie *Systematics* du ZR correspond à la définition que j'ai précédemment donnée pour la taxonomie. J'ai donc considéré que mon corpus d'étude de la taxonomie était constitué des documents indexés avec le terme *Systematics* dans le *Zoological Record*.

Chapitre 8

Mesures des productions et acteurs

8.1 Nombre de publications

Envisager de réaliser un comptage de publications sur une période de temps aussi étendue fait surgir une première question : un article des années 1950 est-il le même objet qu'un article des années 2000 ? Des changements de nombre d'auteur (Baethge, 2008; Medline, 2011), de longueur sont en effet à prévoir. Est-il raisonnable de mettre en comparaison un nombre d'articles longs, écrits en 1950 par un seul auteur, avec un nombre d'articles courts, écrits en 2008 par plusieurs auteurs ? Pour pallier ce biais j'ai choisi de mesurer la distribution au cours du temps du nombre de publications pour la spécialité *Systematics* en la comparant avec quatre autres spécialités : *Ecology OR Habitat*¹, *Conservation*, *Evolution*, *Behaviour*.

8.1.1 Compte de présence

Méthodologie

J'ai réalisé successivement 5 recherches dans le champ SD (Subject Descriptors) du *Zoological Record* :

« SD = "ecology OR habitat" »

« SD = "systematics" »

« SD = "conservation" »

1. Etant donné les définitions proposées par le *ZR* pour *Ecology* et *Habitat* (tableau 7.1), je n'ai pas trouvé judicieux de les séparer.

« SD = "evolution" »

« SD = "behaviour" »

Chaque recherche a ensuite été analysée avec l'outil *Analyze results*. J'ai ainsi obtenu la répartition des articles au cours du temps pour chacune d'entre elles. Les publications sont ici comptées en compte de présence : si un article est interdisciplinaire, il est compté dans chaque discipline. La moyenne mobile sur 3 ans est ensuite calculée pour homogénéiser les résultats. Elle consiste à remplacer chaque valeur par une moyenne ; cette moyenne étant calculée à partir de cette valeur, de la valeur de l'année qui précède et de la valeur de l'année qui suit. Les résultats sont présentés figure 8.1a. J'ai également regardé la part relative de chacune de ces disciplines, i.e. la proportion de documents représentée par chaque discipline au cours du temps. Ce calcul est celui de l'indicateur « part de publications » de l'Observatoire des Sciences et des Techniques. Part de publications (%) = $\frac{\text{Nombre de publications de l'acteur}}{\text{Nombre de publications de la référence}} \times 100$. Les résultats sont présentés figure 8.1b.

Résultats

Les données brutes de la figure 8.1a montrent que le nombre absolu de travaux de systématique publié par an suit une courbe croissante avec le temps, à l'exclusion des deux périodes de guerre et d'une légère diminution après 1985. On observe également une augmentation nette après 1995. La tendance est similaire pour les autres disciplines : écologie, évolution, éthologie.

En revanche, si l'on s'intéresse à la part relative dans la base, c'est-à-dire au volume représenté par les publications de systématique, on constate que celui-ci est en constante diminution : s'il représente 40 % des publications de 1875, il chute jusqu'à 15 % en 1995 (figure 8.1b).

Il ne semble donc pas y avoir de déclin de la discipline à proprement parler. En outre, la diminution observée à partir du milieu des années 1980 est retrouvée au niveau de la distribution globale des références du *ZR* (voir figure 3.5 page 28). Il s'agirait donc, non pas d'une disparition de la systématique, mais d'une augmentation des autres disciplines.

On observe une modification importante des courbes pour toutes les disciplines autour de 1970. Cela ne correspond pas à un événement connu d'intervention sur la base (voir *Historique*, page 18). Il semblerait tout de même que cela soit lié au *ZR* et pas à la littérature : « *We have looked at this and believe that this dip cannot be reliably related to the published literature. It appears that around that time the number of items included in Zoological Record was quite variable year on year and so it is probably just an artefact of staffing levels or availability of literature for*

8.1 Nombre de publications

indexing. Around that time I believe there was also an office relocation which may have affected production capacity for Zoological Record » (Nigel Robinson, comm. pers.).

8.1.2 Compte fractionné

Méthodologie

Les résultats précédents sur le nombre de publications a été obtenu en faisant un comptage direct, c'est-à-dire que si une publication contenait à la fois les mots *Systematics* et *Ecology*, elle été comptabilisée dans les deux disciplines. Or nous avons vu qu'il existe une méthode de comptage prenant en compte l'interdisciplinarité : le comptage fractionné, qui pourrait donner des résultats très différents. Si l'on reprend l'exemple précédent, cette méthode comptera 0,5 pour *Systematics* et 0,5 pour *Ecology*. J'ai cherché à réaliser ce comptage fractionné pour voir si les résultats diffèrent d'avec le comptage direct. Cela nécessitait d'avoir accès, pour chaque document, à tous ses mots d'indexation ; donc à la notice bibliographique complète. L'export des données étant limité à des tranches de 500 articles maximum, je n'ai pu réaliser ce comptage que pour un petit nombre d'années seulement. J'ai choisi les années correspondant aux 5 années définies dans la section [Distribution des articles dans le ZR](#) page 28 : 1929, 1950, 1980, 1998, 2008. L'ensemble du traitement a été réalisé sous **R**. Les résultats en compte de présence sont retrouvés figure 8.2a et les résultats en compte fractionné sont présentés figure 8.2b.

Résultats

Les deux figures nous montrent que l'allure générale des courbes est conservée. Les résultats en compte de présence sont donc validés.

En revanche, les valeurs diffèrent entre les deux courbes. Il n'est pas possible de faire un test de χ^2 ou assimilé pour tester l'homogénéité de ces deux types de comptage. En effet, il faudrait pour cela que les classes soient exclusives, c'est-à-dire qu'un article ne puisse appartenir qu'à l'une ou l'autre des spécialités. Or, c'est justement l'interaction de ces classes que ces résultats mettent en avant. Dans ce cas, comment confirmer l'augmentation de l'interdisciplinarité ? En créant une autre variable à deux classes : "travail interdisciplinaire" ou "travail monodisciplinaire". Le tableau 8.1 présente les proportions de travaux de systématique et de quatre autres disciplines (Écologie, Évolution, Conservation, Éthologie) suivant qu'ils ont été indexés avec un seul ou plusieurs mots-clés disciplinaires.

		1929	1950	1980	1998	2008
Systématique	1 discipline	98.6	99.1	82.9	72.8	72.2
	2 et plus	1.4	0.9	17.1	27.2	27.8
Autres	1 discipline	99.3	94.5	84.2	74	73.3
	2 et plus	0.7	5.5	15.8	26	26.7
χ^2 p-value		$3.51e-2$	$1.96e-7$	$4.9e-3$	$2.2e-2$	$3.1e-2$

TABLEAU 8.1 – Proportion d’interdisciplinarité en Systématique et dans l’ensemble de quatre autres disciplines en 1929, 1950, 1980, 1998, 2008. Test du χ^2 . Autres = Écologie, Évolution, Conservation, Éthologie.

On observe que le nombre d’articles indexés avec plusieurs mots-clés disciplinaires augmente avec le temps, que ce soit en Systématique ou dans les quatre autres disciplines. À partir des années 1980, cette interdisciplinarité est plus importante en Systématique que dans les autres disciplines (test du χ^2 , seuil de 5 %, dernière ligne du tableau 8.1). En 2008, le nombre d’articles indexés à la fois en Systématique et dans une autre discipline atteint 28 %. On assiste donc à une modification des productions de recherche en systématique : les articles ne concernent plus seulement la production de nouveaux taxons. Quelles sont ces modifications ?

Le tableau 8.2 présente la répartition des articles multidisciplinaires incluant la systématique. Ainsi, l’Évolution est associée à Systématique dans 7 % des cas en 1950, et progresse jusqu’à 70 % des articles multidisciplinaires en 2008, soit 19 % de l’ensemble des articles de Systématique. Cela se fait au détriment de l’Écologie, qui passe de 93 % des associations en 1950 à 37 % en 2008. Les associations Systématique - Conservation restent rares : si elles sont de 1,5 % en 1980, elles atteignent 5,4 % seulement en 2008. Notons que le total des colonnes ne fait pas 100 car il peut y avoir des associations multiples (par ex. Systématique - Ecologie - Evolution) et également des associations non prises en compte ici (Systématique - Behaviour par ex.).

DISCIPLINE	1950	1980	1998	2008
Evolution	7	56	64	70
Ecology	93	43	39	37
Conservation	0	1,5	3	5,4

TABLEAU 8.2 – Nature des associations multidisciplinaires incluant la Systématique et trois autres disciplines : Evolution, Ecology et Conservation en 1950, 1980, 1998, 2008. Pourcentages. Source : Zoological Record.

Ces chiffres montrent clairement la place croissante prise par les études sur l’évolution au sein de la Systématique, ainsi que la marginalité des liens existant entre Systématique et Conservation.

Voyons à présent une mesure du nombre d’espèces et sous-espèces décrites depuis 1864.

8.2 Nombre d'espèces et sous-espèces décrites

8.2.1 Méthodologie

Le comptage du nombre d'espèces et sous-espèces décrites par an entre 1864 et 2010 a été réalisé par *Thomson Reuters* en se basant sur les données du *Zoological Record*. Les données sont disponibles dans l'onglet *Metrics* de leur Index to Organism Names² et reproduites figure 8.3a. La figure 8.3b représente la moyenne mobile sur trois ans de ces données.

8.2.2 Résultats

On observe un pic de descriptions autour de 1890, puis une augmentation importante à partir des années 1905 et 1965 (figure 8.3b). Les inflexions observées entre ces deux périodes correspondent aux deux grandes guerres mondiales. On observe ensuite une diminution du nombre d'espèces et sous-espèces décrites à partir de 1985, et une augmentation à partir de 1995. Ces données vont donc dans le même sens que les données de publication.

Voyons maintenant l'évolution du nombre d'auteurs publiant en systématique.

8.3 Nombre de taxonomistes

8.3.1 Méthodologie

J'ai réalisé des requêtes successives dans le *Zoological Record* et utilisé l'outil *Analyze Results* pour estimer le nombre d'auteurs de travaux de systématique publiant chaque année entre 1959 et 2010 ainsi que la proportion représentée par ces auteurs sur l'ensemble des auteurs de zoologie.

Pour obtenir le nombre d'auteurs de systématique, j'ai exécuté la requête

```
SD = "Systematics" AND PY = XXXX
```

pour chaque année. Pour obtenir le nombre d'auteurs décrivant de nouvelles espèces, j'ai exécuté la requête

```
SD = ("New species" OR "Sp nov" OR "New subsp" OR "Subsp nov") AND  
PY = XXXX
```

2. <http://www.organismnames.com>

pour chaque année. En 2008, les documents traitant de nouvelles espèces ou sous-espèces représentent 55 % de l'ensemble des documents de systématique. Pour obtenir le nombre d'auteurs total, j'ai été limitée par l'outil Analyze Results. En effet, celui-ci ne renvoie plus de nombre précis dès que les occurrences comptées dépassent 100 000. Or, à partir de 1998 le nombre d'auteurs est supérieur à 100 000. J'ai donc scindé ma recherche de manière à obtenir les résultats en deux fois. J'ai exécuté, pour chaque année, la requête

$$(IC = "ZOOR*" NOT TS = "Vertebrates") AND PY = XXXX$$

suivie de la requête

$$(IC = "ZOOR*" AND TS = "Vertebrates") AND PY = XXXX$$

Il arrivait que les mentions *ANON*, *ANONYMOUS* ou *ABSTRACTS* remplacent les auteurs. Dans ce cas je les ai ôtées de mon décompte. Les résultats sous forme de moyenne mobile sur 3 ans sont présentés figure 8.4 : partie 8.4a pour le nombre et partie 8.4b pour les proportions.

8.3.2 Résultats

La figure 8.4a nous montre que le nombre d'auteurs ne cesse d'augmenter entre 1959 et 2010 ; et que cette augmentation est plus importante pour les auteurs de systématique à partir de 1998. La figure 8.4b nous montre en revanche que la proportion de ces différents auteurs diminue avec le temps sauf entre 1998 et 2006 pour les auteurs de systématique. Ces données ne sont donc pas en faveur d'une diminution du nombre de taxonomistes à proprement parler. Les diminutions du nombre d'articles et de nouvelles espèces décrites observées entre 1985 et 1998 ne sont pas retrouvées ici. En revanche on retrouve figure 8.4b la chute observée entre 1965 et 1970 pour les publications décomptées en part relatives (figure 8.1b page 96). Il doit s'agir, là encore, d'un événement en lien avec la base de données (voir page 91).

Pour finir, voyons l'évolution du nombre de revues dans ces disciplines.

8.4 Nombre de revues

8.4.1 Méthodologie

J'ai réalisé des requêtes successives dans le *Zoological Record* et utilisé l'outil Analyze Results pour estimer le nombre de revues publiant chaque année des travaux de systématique entre 1978 et 2010 ainsi que le nombre de revues publiant de

8.4 Nombre de revues

nouvelles espèces et sous-espèces. Pour obtenir le nombre de revues de systématique, j'ai exécuté pour chaque année la requête

Subject Descriptors = "Systematics" AND Publication Year = XXXX AND
Document Type = article

Pour obtenir le nombre de revues publiant des descriptions d'espèces et sous-espèces, j'ai exécuté pour chaque année la requête

Subject Descriptors = ("New species" OR "Sp nov" OR "New subsp" OR "Subsp
nov") AND Publication Year = XXXX AND Document Type = article

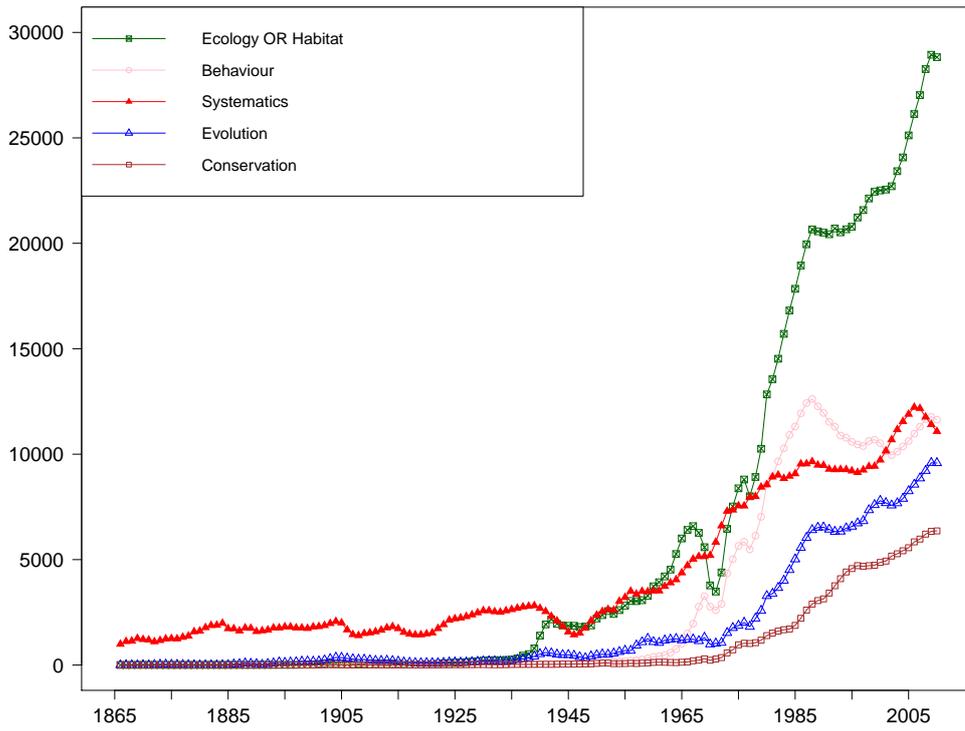
Pour obtenir le nombre total de revues, j'ai exécuté pour chaque année la requête

Identification Code = "ZOOR*" AND Publication Year = XXXX AND Document
Type = article

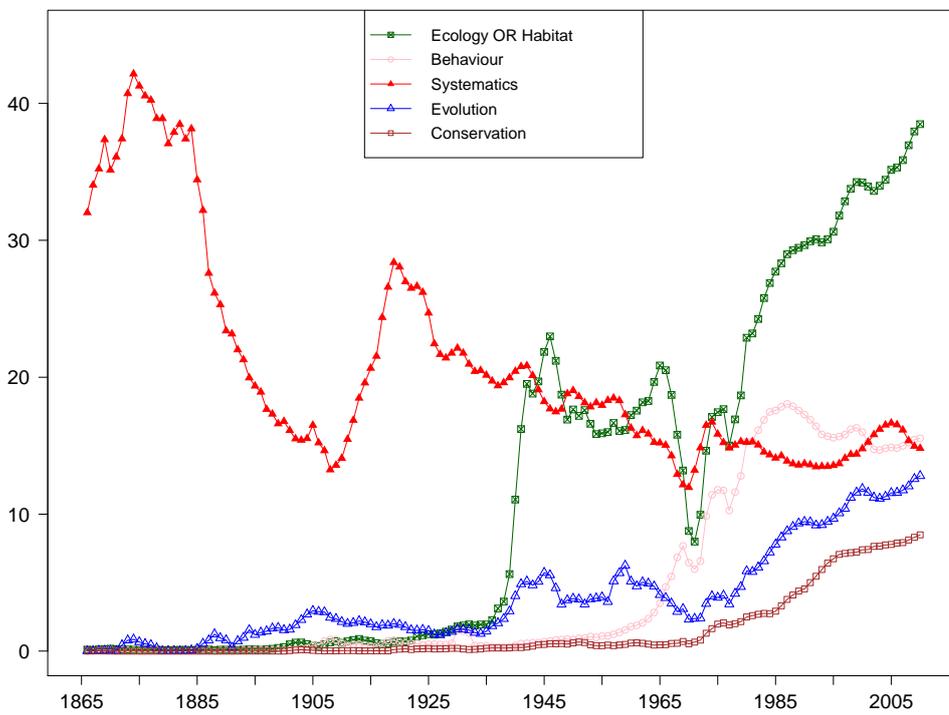
Les résultats sous forme de moyenne mobile sur trois ans sont présentés figure 8.5 : partie 8.5a pour le nombre de revues et partie 8.5b pour les proportions prises par ces revues.

8.4.2 Résultats

Le nombre de revues de systématique a légèrement augmenté, celui publiant de nouvelles espèces et sous-espèces a légèrement diminué (figure 8.5a). Cela peut s'expliquer par une modification du paysage de publication : contrairement à 1978, 2 revues concentrent aujourd'hui un grand nombre de descriptions : *Zootaxa* et *Acta Zootaxonomica Sinica* (voir figure 8.6). En termes de proportion, on observe une diminution du nombre de revues publiant des articles de systématique et publiant de nouvelles espèces ou sous-espèces entre 1980 et 1992. L'étude du nombre de revues publiant des travaux de systématique ne permet donc pas, là encore, de parler de déclin ; en revanche la période de moindre description d'espèces et moindre production de travaux de systématique correspond à une période pendant laquelle on observe une augmentation des revues des autres disciplines.



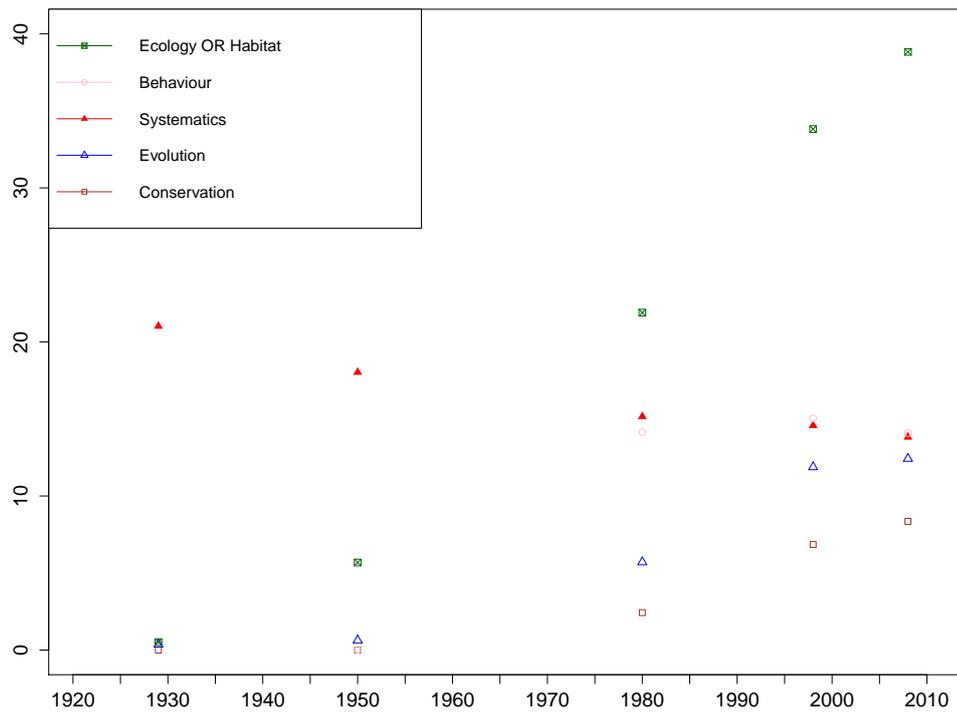
(a) Nombre de publications



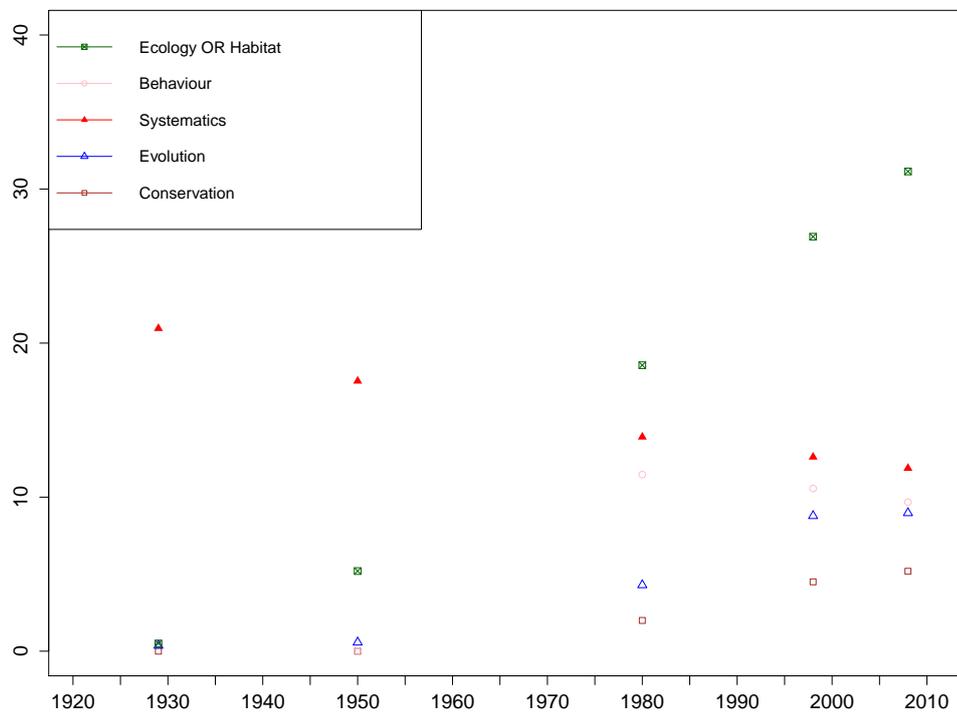
(b) Proportion de publications

FIGURE 8.1 – Nombre et proportion de publications pour 5 spécialités du premier niveau du thésaurus du Zoological Record. Compte de présence, moyenne mobile sur 3 ans.

8.4 Nombre de revues

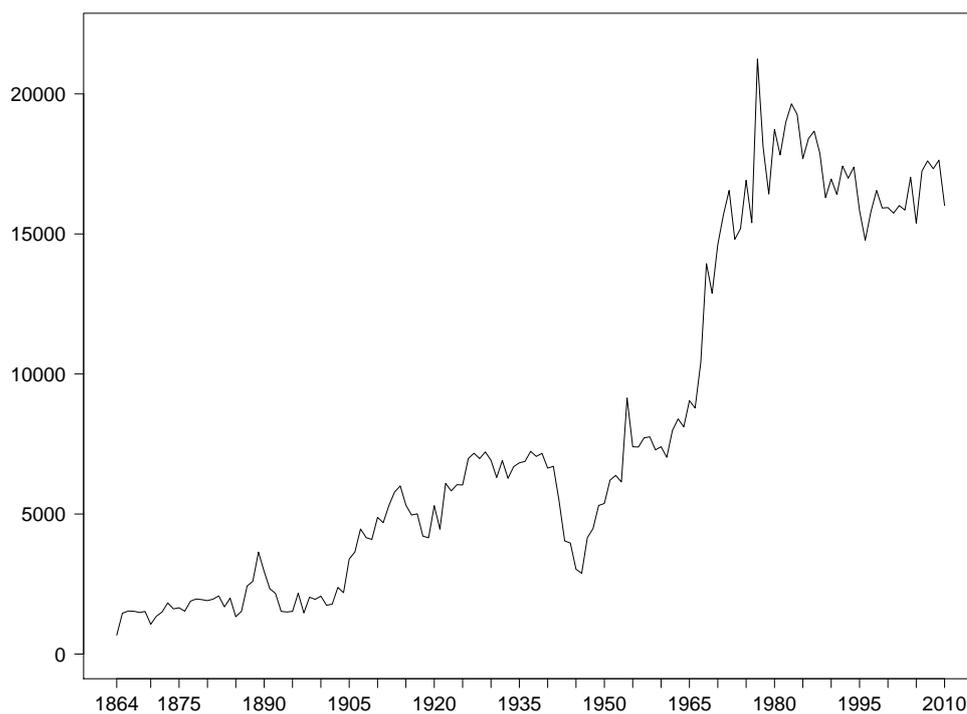


(a) *Compte de présence*

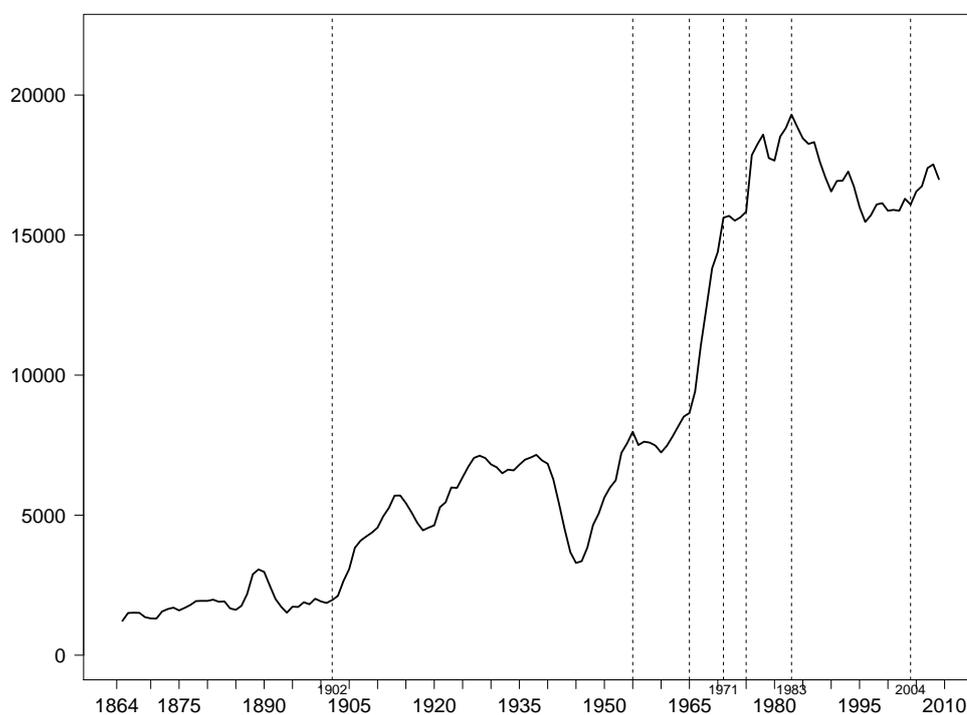


(b) *Compte fractionné*

FIGURE 8.2 – *Proportion des publications zoologiques par spécialité en 1929, 1950, 1980, 1998, 2008. $N_{1929} = 11\,438$ références ; $N_{1950} = 14\,116$; $N_{1980} = 57\,069$; $N_{1998} = 66\,132$; $N_{2008} = 76\,486$.*



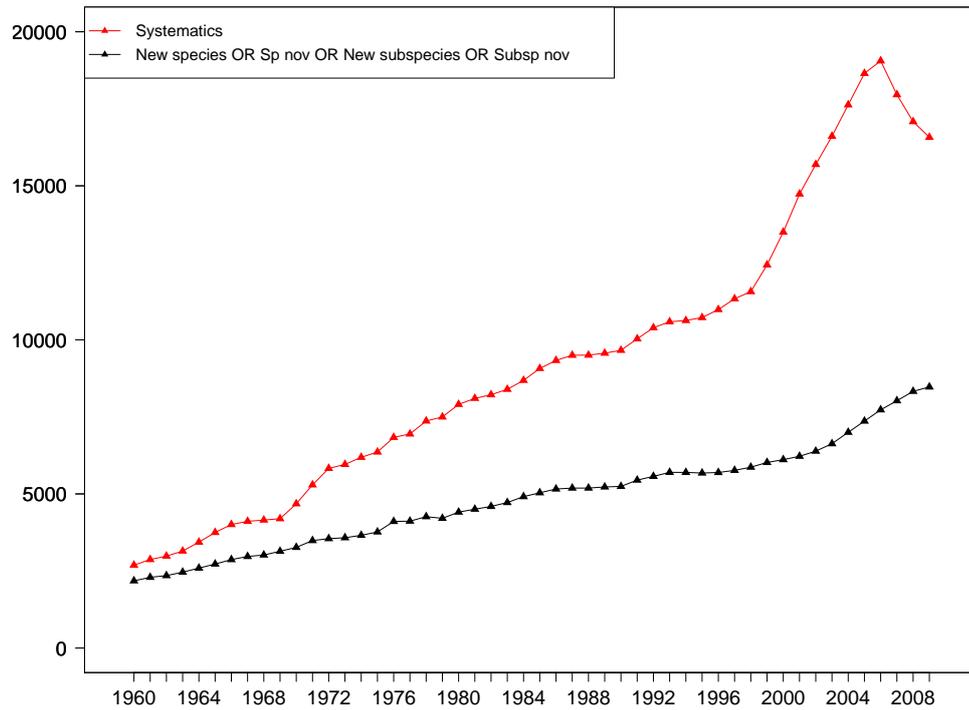
(a) *Données brutes*



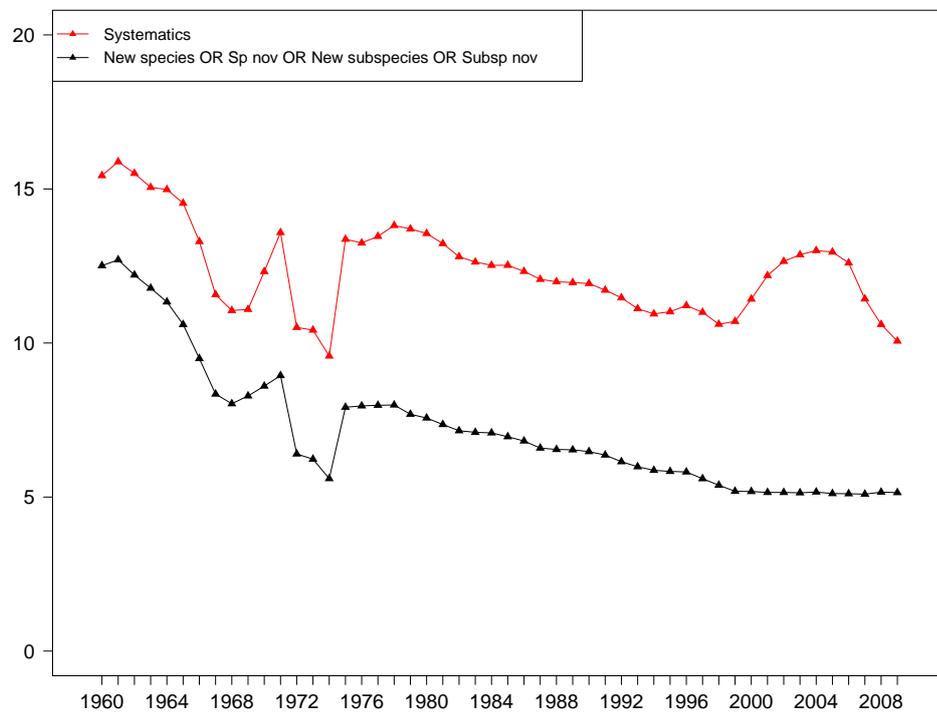
(b) *Moyenne mobile sur 3 ans*

FIGURE 8.3 – Nombre d'espèces et sous-espèces décrites entre 1864 et 2010 en zoologie. Données issues de l'Index to Organism Names issu du Zoological Record. $N = 1\ 182\ 557$ taxons.

8.4 Nombre de revues

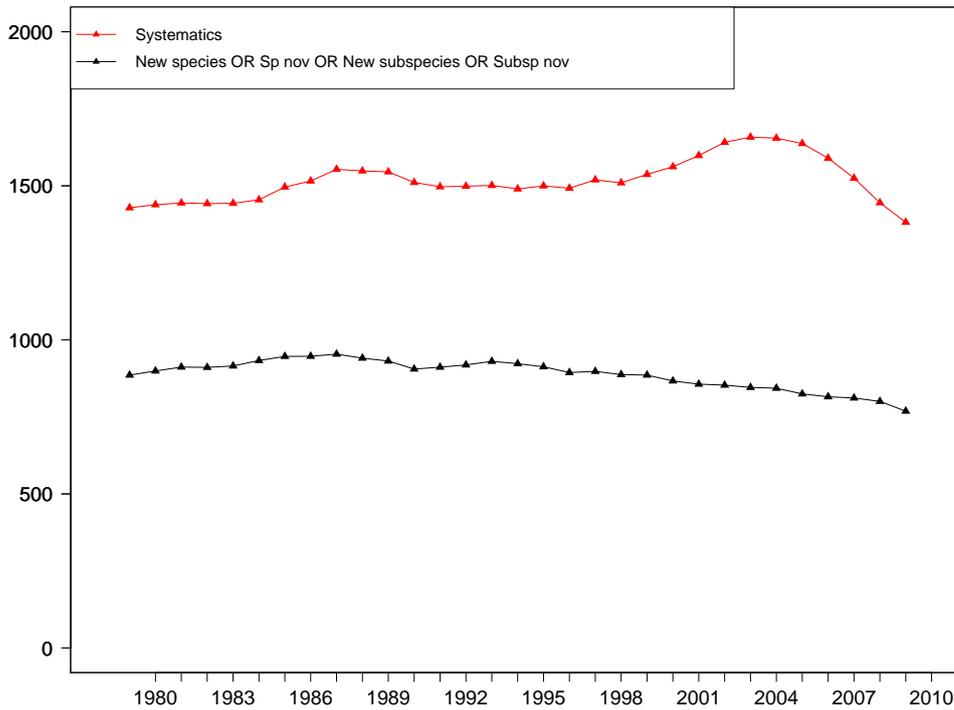


(a) Nombre d'auteurs de travaux de systématique

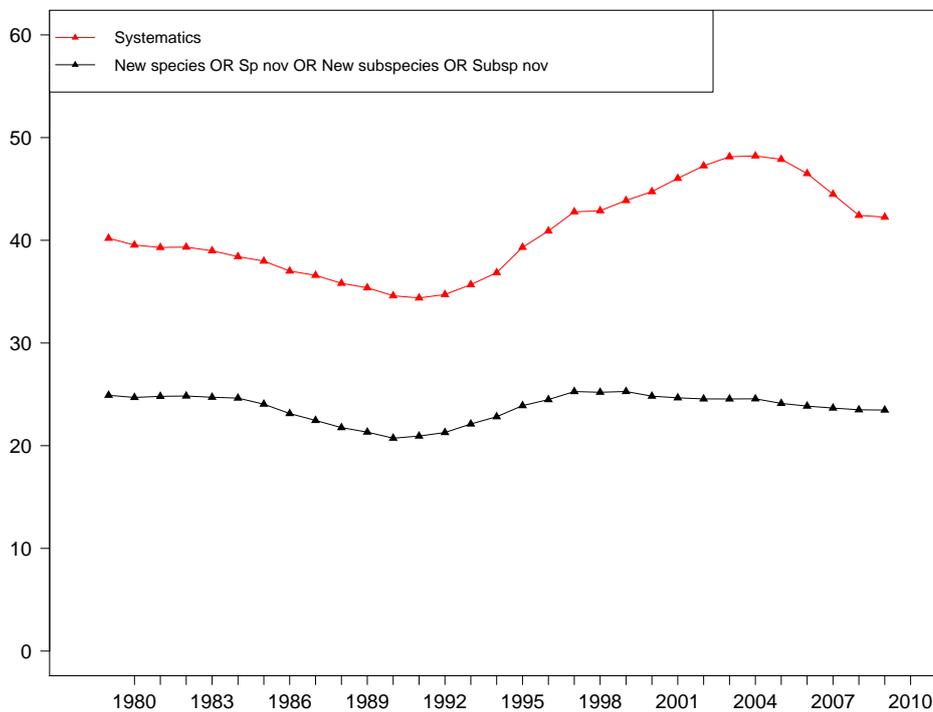


(b) Proportion d'auteurs de travaux de systématique

FIGURE 8.4 – Nombre et proportion d'auteurs de travaux de systématique recensés dans le Zoological Record pour la période 1978-2010. Moyenne mobile sur 3 ans.



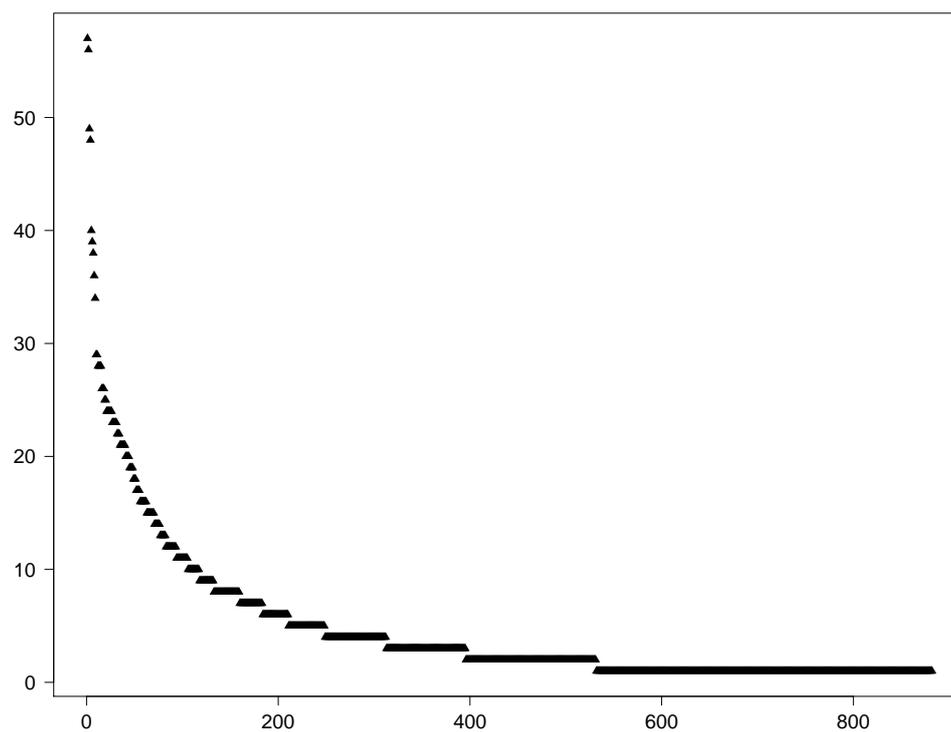
(a) Nombre de revues publiant des articles de systématique



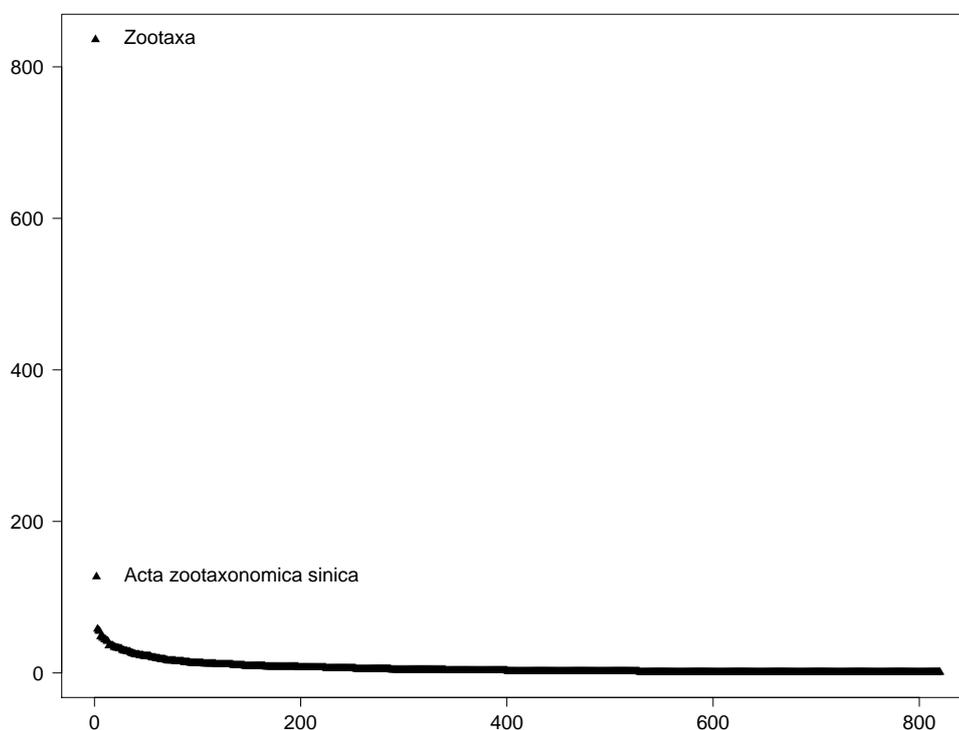
(b) Proportion de revues publiant des articles de systématique

FIGURE 8.5 – Nombre et proportion de revues publiant des articles de systématique et recensées dans le Zoological Record pour la période 1978-2010. Moyenne mobile sur 3 ans.

8.4 Nombre de revues



(a) 1978. $N = 882$ revues



(b) 2008. $N = 820$ revues

FIGURE 8.6 – Nombre d'articles publiant de nouvelles espèces ou sous-espèces, par revue, en 1978 et 2008. Données issues du Zoological Record.

Chapitre 9

Discussion

9.1 Absence de déclin global

Les séries de mesures effectuées dans ce travail ne sont pas indépendantes puisqu'elles proviennent toutes de la même base de données et qu'elles représentent un même objet, la publication scientifique, observé sous des facettes différentes. Néanmoins elles permettent de relativiser l'image caricaturale qui peut être donnée de la taxonomie comme une discipline « sur le point de disparaître ». En outre, le seul déclin vraiment observé (1985-1995) est à questionner dans la mesure où il semble lié à un moindre référencement des publications zoologiques dans le *ZR* (cf. figure 3.5 page 28).

Ces résultats sont confirmés par les travaux de Joppa *et al.* (2011) qui ont étudié les tendances de description d'espèces et d'investissement taxonomique (*taxonomic effort*) de 6 taxons (plantes à fleurs, mollusques du genre *Conus*, araignées, amphibiens, oiseaux et mammifères) entre les années 1800 et 2000 à partir de bases de données indépendantes. Ces auteurs constatent une augmentation exponentielle du nombre d'espèces décrites et de taxonomistes, excepté pour les oiseaux. Dans leur étude, le ratio du nombre d'espèces décrites par taxonomiste décroît, ainsi que le nombre d'auteurs publiant sur plusieurs familles au cours de leur carrière. Ils en tirent la conclusion que les auteurs décrivant des espèces sont de moins en moins cantonnés à la taxonomie et qu'ils sont au contraire de plus en plus spécialisés, taxonomiquement parlant. Cela correspond avec les résultats que j'ai obtenus, à savoir l'interdisciplinarité croissante des travaux de systématique. Il serait intéressant de déterminer si cette polyvalence est le résultat de l'évolution thématique naturelle de la discipline, ou si elle est le résultat des politiques de publication. En effet, le plus haut facteur d'impact des journaux traitant d'évolution est 14,448 (*Trends in Ecology & Evolution*); 15,253 pour l'écologie (*Ecology Letters*) et 6,346 pour la conservation

(*Global Change Biology*) alors qu'il est de 4,800 pour les journaux de zoologie (*Wildlife Monographs*)¹. Étant donné l'importance du facteur d'impact dans l'évaluation de la recherche aujourd'hui, quelle que soit la discipline (Thomaz *et al.*, 2011; Ortnner, 2011; Esposito, 2011), il est légitime de se demander si l'élaboration de travaux multidisciplinaires ne serait pas dans une certaine mesure un moyen de publier dans des revues à facteur d'impact plus élevé que ce à quoi l'on pourrait prétendre en restant dans sa discipline. Cette question de l'influence des politiques d'évaluation sur les pratiques de recherche est toutefois une démarche complexe qui ne peut être réduite à l'étude d'un seul indicateur.

Une étude bibliométrique de Liu *et al.* (2011) a montré que les études sur la biodiversité ont augmenté de manière exponentielle à partir du milieu des années 1990 ; années qui ont suivi l'établissement de la Convention sur la diversité biologique à la conférence de Rio de Janeiro (Anonymous, 1992). On peut donc penser que l'augmentation du nombre de publications de systématique et du nombre d'espèces et sous-espèces décrites à partir de 1995 est également une conséquence de ce sommet.

9.2 Vers une redistribution des forces ?

Nous avons vu qu'il n'est pas possible de parler de déclin à l'échelle internationale. Toutefois, n'y aurait-il pas *un* déclin, mais *des* déclins, multiples et localisés ? Il serait possible que certains pays voient leur population de taxonomistes décroître, ou que certains taxons soient moins étudiés aujourd'hui qu'auparavant.

Je me suis intéressée à la répartition des travaux de systématique suivant les continents (figure 9.1a) et les pays (figure 9.1b) des seuls auteurs dont l'adresse est mentionnée dans le *ZR* (i.e. l'auteur de correspondance), en 1998 et 2008. Je me suis également intéressée à la répartition des travaux de systématique suivant les taxons pendant ces mêmes années. Les résultats sont présentés tableau 9.1.

On remarque figure 9.1b une régression des pays leaders, et principalement : Royaume-Uni, Russie, Australia ; au profit de pays émergents : Chine principalement, mais également Argentine. Or, il s'avère que le Royaume-Uni et l'Australie sont les pays qui ont mis en place un suivi de leur effort taxonomique et dont les rapports (Australian Government department of the Environment Water Heritage and the Arts, 2006; House of Lords Science and Technology Committee, 2008) sont fréquemment cités pour argumenter le déclin.

En outre, si l'on s'intéresse aux pays des 32 auteurs qui ont pris part à la discussion *Why Taxonomy does NOT matter ?* posée fin avril 2011 sur la liste de diffusion

1. Données issues du *Journal Citation Report* de 2010 pour les catégories *Evolution, Ecology, Biodiversity Conservation, Zoology*.

9.2 Vers une redistribution des forces ?

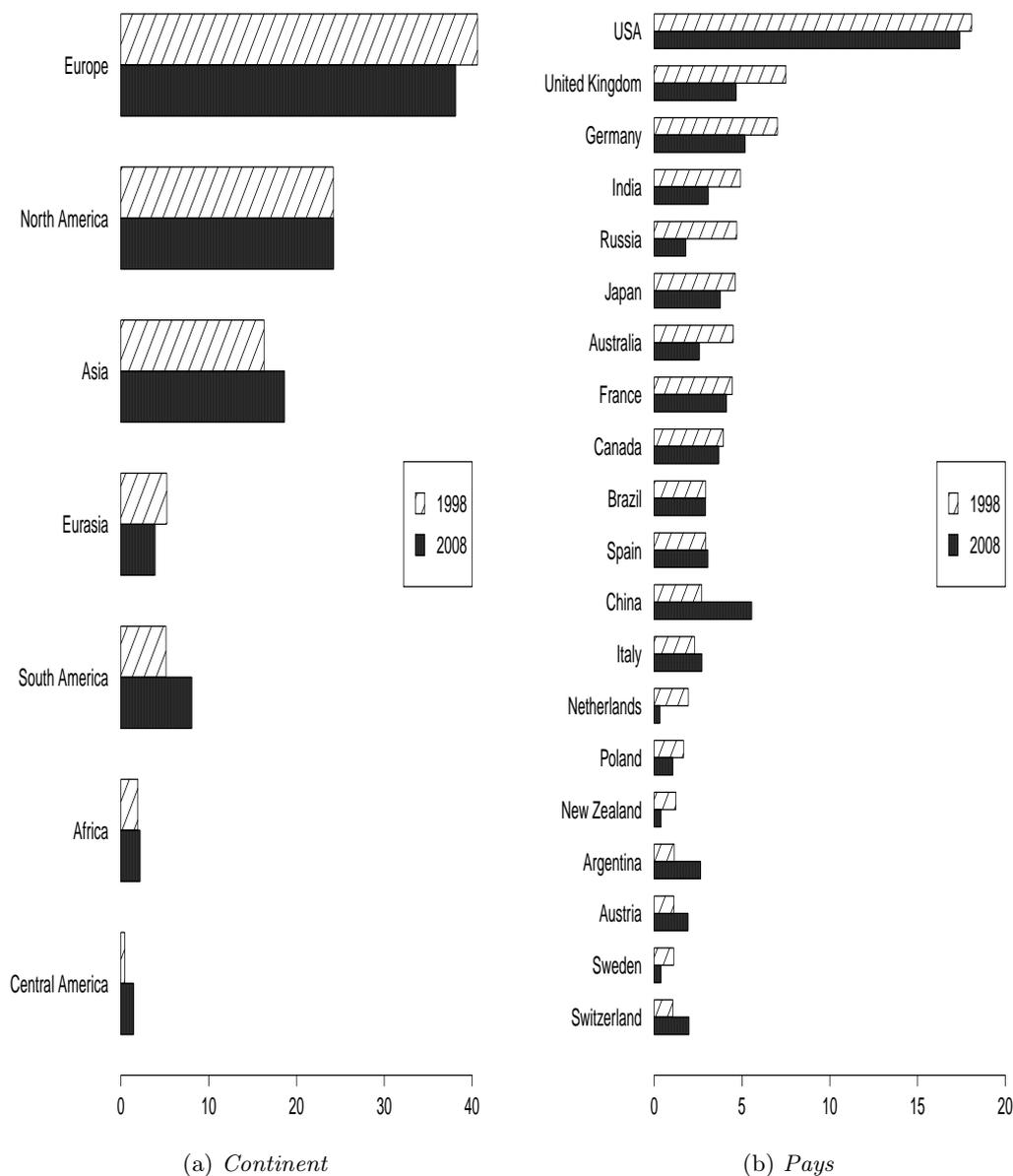


FIGURE 9.1 – Proportion des travaux de systématique par continent et pour les 20 premiers pays de l'auteur de correspondance en 1998 et 2008. Données issues du Zoological Record. $N_{1998} = 9\ 642$ dont 4,4 % sans adresse et $N_{2008} = 10\ 575$ références dont 3,3 % sans adresse. Les scripts d'extraction des pays et d'ajout des continents ont été réalisés sous **R**. Les continents ont été attribués à chaque pays en suivant les données du World Factbook (Central Intelligence Agency, 2008). Eurasia = Moyen-Orient.

Taxacom² et mentionnée en introduction, on observe qu'un tiers d'entre eux sont situés aux États-Unis, 22 % en Australie et... 12 % au Royaume-Uni.

2. <http://mailman.nhm.ku.edu/pipermail/taxacom/2011-April/072162.html> par le taxoniste allemand Fabian Haas.

TAXON	1998	2008	ACCROISSEMENT (%)
INVERTEBRATES	7170	8043	12
ARTHROPODS	5303	6090	15
INSECTS	4088	4575	12
CHORDATES	2323	2457	19
VERTEBRATES	2295	2731	19
BEEETLES	1263	1501	19
FISH	682	932	37
MOLLUSCS	788	853	8
CHELICERATES	514	789	54
ARACHNIDS	505	781	55
TRUE FLIES	649	735	13
LEPIDOPTERANS	761	687	-10
MAMMALS	689	686	0
CRUSTACEANS	639	626	-2
HYMENOPTERANS	561	598	7
REPTILES	484	553	14

TABLEAU 9.1 – *Accroissement du nombre de travaux de systématique entre 1998 et 2008 pour les taxons totalisant plus de 5 % des publications référencées dans le Zoological Record en 2008. Source : Zoological Record. Résultats issus de l'outil Analyze results by Taxa Notes du ZR. Requête : « SD = "Systematics" AND PY = 1998 ». Les taxons sont issus des catégories du Zoological Record et sont conservés même s'ils se recouvrent (par ex. : Invertebrates et Insects). Chiffres d'accroissement arrondis à la valeur supérieure.*

Il serait donc possible qu'il y ait un déclin dans certains pays développés, contrebalancé par un accroissement dans les pays en émergence (Tancoigne *et al.*, 2011). Cette thèse est également évoquée par Joppa *et al.* mais nécessiterait des analyses plus fines pour être validée.

En ce qui concerne les taxons, on observe tableau 9.1 que l'accroissement du nombre de travaux n'est pas le même pour tous les taxons. Il est très important pour les Chélicérates (araignées...), et en particulier pour les Arachnides (55 %), ainsi que pour les poissons (37 %). En revanche il est négatif ou nul pour les Lépidoptères (papillons, -10 %), les Crustacés (-2 %) et les mammifères (nul).

Toutefois, si l'on fait le même calcul entre 1980 et 2008, on observe un accroissement pour tous les taxons (non montré). Aussi, si l'hypothèse d'un déclin ciblé sur des taxons donnés est probable, elle nécessiterait d'être confirmée.

Pour conclure, dans la mesure où il n'est pas possible de parler de déclin global de la discipline, mais plutôt d'un déplacement des forces présentes en taxonomie, je propose de regarder de plus près les travaux portant sur le déclin de la discipline. Quelle en est l'origine ? À quand remonte la naissance de cette inquiétude ? Quels ont été les arguments mis en avant, les solutions proposées ?

Quatrième partie

Le déclin, une inquiétude récurrente

Chapitre 10

Une inquiétude récurrente dans *Nature* et *Science* depuis 1946

J'ai choisi comme première approche d'analyser les discussions portant sur la taxonomie au sein de deux grands journaux généralistes : *Nature* et *Science*. Le chapitre suivant présente le résultat de cette analyse. J'ai identifié leur fréquence d'apparition ainsi que les thématiques abordées autour de cette discipline. Les questions qui m'ont guidée étaient les suivantes : (1) existe-t-il des périodes-clés de débat autour de la discipline ? (2) Si oui, quelle est l'importance prise par cette idée de déclin au sein de ces débats ?

10.1 Méthodologie

J'ai exécuté la requête

Title = ("taxonomy") AND Publication Name = (Nature OR Science)

dans le *Web of Science*. La recherche a volontairement été restreinte aux articles mentionnant le terme *taxonomy* dans leur titre. Une recherche étendue aux mots-clés et résumés donne des résultats similaires. J'ai ensuite utilisé l'outil *Analyze results by Publication Years* pour connaître leur répartition dans le temps (N = 212 documents). La figure 10.1 représente cette répartition temporelle. Parmi ces publications, 158 proviennent de *Nature* et 75 de *Science*. Ce sont principalement des lettres (80 publications), des articles (74) et des éditoriaux (49).

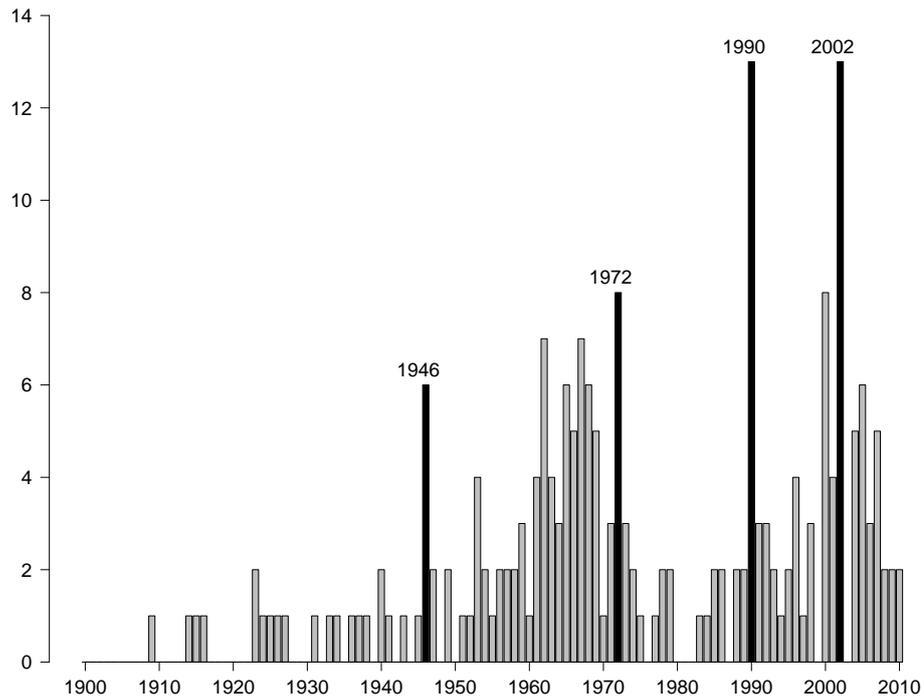


FIGURE 10.1 – Nombre de documents comportant dans leur titre le mot *taxonomy* dans les journaux *Nature* et *Science* entre 1900 et 2010.

10.2 Résultats

On distingue quatre années où le mot a été particulièrement employé dans ces deux revues tout en se distinguant des années avoisinantes : 1946, 1972, 1990 et 2002 (en noir sur le graphique). On observe également un intérêt pour la taxonomie à la fin des années 1960 et au début des années 2000. À quoi correspondent ces différentes périodes ?

10.2.1 L'après-guerre

Suite à la seconde guerre mondiale, Fosberg et Diehl demandent à ce qu'il soit fait un bilan des pertes matérielles engendrées par la guerre pour les collections naturalistes. Au cours de leur argumentation, les auteurs mentionnent brièvement que la « *Systematic botany has traditionally been ill-supported* ». Cette idée est alors reprise et développée dans les publications suivantes. Ainsi, Ball (1946) pense que cela s'explique par le fait que ses résultats (les noms) ne sont pas stables ; Cockerell (1946) que c'est une discipline compliquée, qui se pratique lentement ; et Metcalf (1946) que c'est une discipline qui ne bénéficie pas de la sympathie des autres biologistes. Les deux autres éditoriaux, publiés anonymement dans *Nature*, rappellent la nécessité d'utiliser un ensemble d'approches différentes (Anonymous, 1946a) et d'embaucher

10.2 Résultats

des taxonomistes (Anonymous, 1946b) pour réaliser le travail de compréhension de la faune et de la flore.

Dans tous ces documents, les auteurs justifient l'attention qu'ils portent à cette discipline par l'idée qu'elle est au service des autres sciences :

to place before the other workers in biology as full and complete a record of the forms living in the world [...] (Metcalf, 1946)

Sound taxonomy is the mother of sound ecology (Cockerell, 1946)

Taxonomy has no required standards of preparation or fitness, yet its products may help or harm many (Ball, 1946)

Classification and the identification of material is also fundamental to all other types of biological research (Fosberg et Diehl, 1946)

this emphasis on the importance of taxonomy was not made on account of personal preference but from a sense of responsibility and a desire to serve the interests of biologists as a whole (Anonymous, 1946b)

Une première période de réflexion se trouve donc au sortir de la guerre. Bien que vaguement énoncée, l'idée que la discipline manque de considération fait consensus : « *the real reason for the neglect of taxonomy* » (Cockerell, 1946), « *My earnest plea is for support for taxonomy from all biologists* » (Metcalf, 1946). Elle est perçue comme une discipline au service des autres branches de la biologie, discipline dont le devenir pose question. Cela se pose à la fois en termes de survie des collections (supposées endommagées par la guerre), et en termes d'image (supposée mauvaise et à l'origine d'un manque d'intérêt général). Reproche est fait aux taxonomistes de ne pas faire en sorte d'être compris, et aux autres biologistes de ne pas faire en sorte de les comprendre.

10.2.2 Émergence de nouvelles techniques en taxonomie

Les articles observés en nombre plus important dans les années 1960 font référence à l'utilisation de nouvelles techniques issues de la chimie, de la biochimie et de la taxonomie numérique, pour obtenir des résultats taxonomiques portant essentiellement sur les plantes, les primates et les unicellulaires.

10.2.3 Remise en cause des objectifs

En 1972, la publication de la nécrologie de l'entomologiste A. W. Stelfox entraîna la publication de deux lettres dénonçant le manque de moyens humains et d'entretien

des collections dont souffre la discipline : d'abord en entomologie (Graham, 1972), puis en ornithologie (Bourne, 1972).

Peu de temps avant, un article de Raven *et al.* (1971) avait soulevé un problème nouveau : il ne sera pas possible de décrire à temps toutes les espèces, compte tenu du fait qu'elles s'éteignent à un taux d'extinction que l'on pressent élevé. Face à ce constat, les auteurs remettent en cause la nécessité de tout décrire et proposent une solution technique globale, informatique, pour accélérer le travail taxonomique. Les nombreux articles observés en 1972 correspondent donc à une remise en cause de cette assertion selon laquelle il ne serait pas utile de tout décrire (Wilson, 1972; Evans, 1972; Van der Spoel *et al.*, 1972).

On retrouve les arguments utilitaires précédemment évoqués (Graham, 1972; Evans, 1972), mais on observe également l'apparition d'arguments défendant la taxonomie pour elle-même :

Modern taxonomy is no longer the science of "giving names and nomenclature"; the final result of taxonomy will not be a list of all organisms, with their names, pictures, and full description. The final result has to be the understanding of the species concept, the genus concept, the family concept, and so forth, and the understanding of the relation and interaction between species or between genera (Van der Spoel et al., 1972)

record information about diversity on a continental or global scale, [...] the exploration of the remainder of life on Earth is not only justified but one of the most exciting and potentially fruitful tasks lying ahead (Wilson, 1972)

Au début des années 1970, on voit donc qu'autour d'un événement ponctuel, le décès d'un spécialiste, s'articule la dénonciation d'un manque de moyens pour les collections et d'un manque de taxonomistes. Ce manque est rendu plus aigu qu'à la période précédente en raison de la prise de conscience de l'extinction des espèces. La remise en cause des objectifs et des pratiques de la discipline est lancée. Elle s'accompagne, en réaction, d'arguments insistant sur l'aspect scientifique, et non sur l'aspect de service, de la discipline.

10.2.4 Remise en cause des pratiques

Les 13 articles de l'année 1990 défendent tous l'idée que la taxonomie a besoin de plus de ressources. Ces échanges font suite à deux événements : la suppression de postes au NHM de Londres (Gee, 1990b,a) et un exemple emblématique de problème de conservation dû à une méconnaissance taxinomique (Daugherty *et al.*, 1990).

10.2 Résultats

Ces deux événements sont l'occasion pour certains auteurs de remettre en cause les pratiques de la discipline : Clifford *et al.* (1990) proposent que moins de spécimens soient conservés en collection, Macpherson (1990) qu'elle stabilise ses noms pour redevenir crédible. May (1990) propose une nouvelle mesure de la diversité taxonomique pour les programmes de conservation (*taxonomic distinctness*). Gee (1990c) propose l'utilisation de bases de données informatisées comme solution pour maintenir à jour les changements taxinomiques. L'idée de Clifford *et al.* suscite de nombreuses réponses (Jarzembowski, 1990; West et Conn, 1990; Stirton *et al.*, 1990; Stevens, 1990; Harley, 1990; Chalmers *et al.*, 1990).

La taxonomie est à nouveau défendue comme une discipline au service des autres sciences et de la société. Le discours défendant la taxonomie comme une science autonome, et pas une discipline au service des autres sciences, est minoritaire (Bremer *et al.*, 1990; West et Conn, 1990).

Accurate identification of taxa is of prime importance to most biological research (Clifford et al., 1990)

Taxonomies are not irrelevant abstractions, but the essential foundations of conservation practices (Daugherty et al., 1990)

Taxonomic research [...] is the "bedrock" of all biological research (Gee, 1990a)

Science and industry rely on taxonomists (Harley, 1990)

Without taxonomy to give shape to the bricks, and systematics to tell us how to put them together, the house of biological science is a meaningless jungle (May, 1990)

the naming service is probably one of the most important taxonomic services provided to farmers, doctors, hospitals, [...] (Stirton et al., 1990)

Comme précédemment, on voit s'articuler autour de faits ponctuels, considérés comme emblématiques, un échange d'idées portant sur une remise en question de la discipline. La question posée est toujours la même : s'agit-il de révolutionner sa pratique, ou de lui accorder plus de moyens ?

10.2.5 L'ère informatique

Les publications de 2002 font suite au développement de grandes bases de données en biologie et aux applications qui pourraient en découler pour sortir la taxonomie de la crise dans laquelle les auteurs s'accordent à la trouver.

Pour faire face à la crise que traverse la discipline, Godfray (2002a) invite à mettre en place une taxonomie unitaire, c'est-à-dire une « taxonomie participative », en ligne : « *This draft first web revision would be placed on the web for comments from the community, then after changes have been made in response, it would become the unitary taxonomy of the group* » .

Cette idée de centralisation des ressources, également défendue par *Nature* (Anonymous, 2002) est reprise par Gewin (2002); Bisby *et al.* (2002); Blackmore (2002). Anonymous; Tautz *et al.* surenchérisent sur cette proposition de changement en ajoutant qu'il devra se faire en faisant une part plus importante à la génétique.

A contrario, d'autres auteurs s'élèvent pour dire que cette informatisation ne sera pas suffisante. Ainsi, Oren et Stackebrandt (2002) expliquent qu'elle est déjà bien développée dans le cas des procaryotes, mais insuffisante pour faire progresser la discipline plus avant. D'autres auteurs mettent le doigt sur le risque d'un divorce entre les deux faces de la taxonomie : la fonction de service - rattacher des informations biologiques à un nom - et la fonction de recherche - établir des hypothèses. Il est demandé de faire attention à ce que les noms ne soient pas uniquement vus comme des clés d'accès à l'information biologique, mais également comme des hypothèses scientifiques (Thiele et Yeates, 2002; Knapp *et al.*, 2002).

Un autre débat concerne l'évaluation scientométrique de la taxonomie via les facteurs d'impact (voir Krell, 2002).

10.2.6 L'ère du barcode

Les articles observés en nombre plus important après 2004 correspondent à des articles sur le 300^e anniversaire de la naissance de Linné, ainsi que sur un débat concernant l'utilisation d'une nouvelle technique, le barcode. Cette technique consiste à chercher des marqueurs génétiques qui permettraient d'individualiser, donc de reconnaître, chaque espèce vivante à partir de son ADN.

10.2.7 Conclusion

Cette analyse des articles de *Nature* et *Science* nous montre que la question de la mauvaise santé de la taxonomie est un thème présent régulièrement depuis la seconde guerre mondiale. Il est associé à des remises en causes successives des pratiques de la taxonomie qui s'accompagnent de débats importants. La place qui lui est donnée dans ces journaux témoigne à mon sens de l'importance qui lui est reconnue (majorée par le fait que la requête utilisée, volontairement minimaliste, a laissé de côté tous les articles de taxonomie qui ne mentionnent pas ce terme).

10.2 Résultats

Toutefois, ces revues ne semblent pas être le lieu privilégié d'argumentation du déclin supposé de la discipline. En effet, ces articles sont le plus souvent associés à des événements symboliques : inquiétude portant sur une collection donnée, décès d'un expert, etc.

J'ai donc effectué en parallèle une recherche bibliographique large, sur une période de temps étendue, afin de voir comment le thème du déclin (ou de la bonne santé) de la taxonomie a été argumenté au cours du temps.

Chapitre 11

Des mesures depuis le début du XX^e siècle

Le sujet de l'état et des perspectives de la taxonomie a été abordé dans des contextes variés et par des acteurs très différents. La constitution du corpus de documents traitant de ce sujet était donc une première étape que cette hétérogénéité des sources, du vocabulaire, et, potentiellement, des langues employées, rendait très délicate.

Les documents traitant de ce sujet peuvent en effet être :

- des articles, notes ou lettres éditoriales écrits spontanément par des scientifiques (taxonomistes ou non) ;
- des articles proposés pour une occasion particulière (commémoration, symposium, ouvrage récapitulatif sur les avancées de la discipline) ;
- des discours de présidents de sociétés savantes, prononcés à l'occasion de l'ouverture ou de la clôture des congrès annuels de leur société ;
- des introductions d'ouvrages généralistes sur la discipline ;
- des rapports, commandés par des institutions savantes, des sociétés savantes ou des gouvernements ;
- et, pour finir, des articles écrits par des journalistes, destinés au grand public.

Le vocabulaire que revêt cette thématique est également très varié. Que ce soit pour désigner la discipline elle-même (« histoire naturelle », « zoologie », « taxonomie », « systématique », « ornithologie », « mammalogie », « entomologie »...), le travail consistant à faire un bilan de ses avancées et de son état (« perspectives », « futur », « portrait », « état des lieux »), ou bien encore sa mise en péril (lorsque certains vont parler de « disparition » des « naturalistes », d'autres utiliseront les termes de « crise » en « systématique », de « problèmes » de la « zoologie » ou encore du « déclin » de la « taxonomie »).

S'il était impossible de parvenir à constituer un corpus exhaustif des documents traitant de ce sujet, l'utilisation de différentes stratégies de recherche a néanmoins permis de prendre en compte la double hétérogénéité des sources et du vocabulaire, et ainsi de réunir plus de 200 documents. Le détail de ces recherches est donné pour information en annexe H.

11.1 Une préoccupation dès le début du XX^e siècle

Cette recherche m'a permis de découvrir que l'inquiétude portant sur la santé de la taxinomie est antérieure à la 2^{de} guerre mondiale, et particulièrement en taxonomie botanique. À titre d'exemple, le botaniste nord-américain Hitchcock écrivait en 1916 : « *Toward the end of the century taxonomy not only lost its dominance, but in this country at least [USA] was relegated to an inferior place* » (Hitchcock, 1916). On trouve également des articles plus anciens, centrés sur la disparition, dans l'enseignement, des disciplines d'histoire naturelle : « *When it became fashionable to study physiology, histology, and embryology, the study of systematic natural history was not only neglected, but disappeared from the college curriculum, and the race of naturalists became nearly extinct* » (Merriam, 1893). Ce discours, issu de l'analyse personnelle de leurs auteurs, se retrouve ensuite de façon récurrente tout au long du XX^e siècle, que ce soit dans *Nature* et *Science* ou dans des revues de notoriété moindre. Pour la période précédant la 2^{de} guerre mondiale, on peut voir à titre d'exemple Shear et Clements (1926); Hitchcock (1928); Tansley (1939).

Des mesures apparaissent également très tôt. Ainsi, en 1939, l'entomologiste Brues réalise pour la première fois une mesure du nombre de taxons décrits à l'aide du *Zoological Record* : « *I have been able to collate with the aid of the "Zoological Record" a summary of the new names proposed annually by zoologists over a period of somewhat more than half a century [...] The change in rate over the whole period is consistent, but not great and does not bear out the opinion that is frequently voiced by many zoologists that systematics has gone wild with the needless multiplication of generic names* » (Brues, 1939). Brues tire de son étude la conclusion que la taxonomie est une discipline en pleine croissance qui a su tirer parti des changements opérés en biologie.

Puis en 1959, c'est au tour de l'entomologiste Blackwelder de poser un certain nombre de questions : « *What is a taxonomist? How many taxonomists are there? Where do they live and work? Are they professionals or amateurs? Are they individually adequately trained for their work? Do they altogether deal with all groups of animals throughout the world? Are all these groups well known? Is the work finished on any of them? Is taxonomy still needed? Are there enough taxonomists for the job? Is there a large or a small volume of contributions being made at this period?*

11.2 Des arguments récurrents

Are there ways in which taxonomy can be improved ? » (Blackwelder, 1959). Il réalise sur deux ans un inventaire des taxonomistes à l'échelle mondiale avec l'aide de la *Society of Systematic Zoology* et de l'*International Bureau for Plant Taxonomy and Nomenclature*. Étonné, il constate que le nombre total de taxonomistes est deux fois plus élevé que ce qui était attendu. Il conclut que la discipline est en pleine croissance et que son avenir est prometteur sous réserve de quelques modifications de ses pratiques.

Steyskal fait en 1965 un travail similaire à celui de Brues, dans une autre perspective : celle de déterminer si la taxonomie est sur le point d'avoir décrit toutes les espèces ou bien s'il lui reste encore beaucoup d'espèces à décrire. Il conclut que la taxonomie a encore beaucoup de travail devant elle. Par la suite, la distribution du nombre d'espèces décrites servira à estimer le nombre d'espèces total d'un groupe (voir par exemple White, 1975).

Nous voyons ainsi que lorsque des études quantitatives ont eu lieu à une échelle internationale, elles vont à l'encontre des idées communément admises et surprennent positivement leurs auteurs.

Les états des lieux de la discipline à une échelle nationale débutent dans les années 1960 et seront régulièrement publiés dans le monde (voir par exemple Balcells, 1963; Ride et McCusker, 1978; Advisory Board for the Research Councils, 1979; Edwards *et al.*, 1985). Le suivi à l'intérieur d'un pays est rare à l'exception du Royaume-Uni et de l'Australie, mentionnés dans le chapitre précédent.

11.2 Des arguments récurrents

Au cours de l'analyse de cette littérature, j'ai pu constater qu'un certain nombre d'arguments revenaient fréquemment. Le principal argument rencontré, duquel découle tous les autres, est que la taxonomie est en déclin parce qu'elle souffre d'une mauvaise réputation. Les raisons supposées de cette mauvaise réputation sont nombreuses :

- elle manque de rigueur dans sa méthodologie, produisant des résultats de piètre qualité qui se traduisent en une longue liste de probables synonymies nomenclaturales ;
- elle n'attire plus de chercheurs car elle est beaucoup moins technicisée que les autres disciplines ;
- elle ne permet pas de faire une carrière scientifique digne de ce nom ;
- c'est une discipline obscure et aride ;
- ...

La thèse que j'ai pu observer comme étant la plus répandue est que la taxonomie souffre du développement des autres disciplines ; particulièrement si celles-ci sont liées à l'élaboration de nouvelles techniques.

every opportunity should be taken to point out to him [a student] the objects of fascinating interest that are found beside the way. Failure to do this leads to the production of those near-sighted naturalists, who, in the expressive words of Professor Forbes, "must have nature boiled in corrosive sublimate solution and fried in paraffine and sliced by a microtome before they care for it" (Weed, 1894).

[Taxonomy and associated research in organismal biology dropping at a relatively fast rate] *is likely due to the rise of molecular studies and the biotechnology industry that gives options for scientists (and probably non-naturalists) to choose a more "profitable" field of study (Loebl et Leschen, 2005).*

C'est volontairement que je n'ai cité ici que deux extraits de ces documents mettant en cause le développement technique de la biologie au détriment de l'histoire naturelle, ou de la taxonomie. J'ai choisi deux citations très éloignées dans le temps pour montrer que cette idée est vraiment ancienne et qu'elle perdure jusqu'à aujourd'hui. Une question surgit alors : ce discours ne serait-il pas inhérent au développement de la biologie, accompagnant chaque nouvelle spécialisation de la discipline ?

En parallèle de ce discours se trouve un autre point de vue, portant sur ce que *devrait être* la taxonomie : une discipline synthétisant toutes les données de la biologie pour construire une grande synthèse de nos connaissances en biologie :

the ideal of taxonomy is the utilization of the results obtained by all the branches of botany; it is the expression of the sum of the knowledge to which all contribute; it is the philosophy of botany in that it correlates the parts into a harmonious and ever growing whole (Hitchcock, 1916).

Biology, the life science, the understanding of the development, adaptation, variability, and diversity of the most natural and original of all matter, is not dependent on, but culminates in, the science of taxonomy (Van der Spoel et al., 1972).

Cet argument du "taxonomiste complet" (*complete taxonomist*, Hatch, 1948) est celui qui a été repris par Dayrat (2005) sous le terme de *taxonomie intégrative* : « *'Integrative taxonomy' is defined as the science that aims to delimit the units of*

11.2 Des arguments récurrents

life's diversity from multiple and complementary perspectives (phylogeography, comparative morphology, population genetics, ecology, development, behaviour, etc.) »

À la lumière de cette recherche bibliographique nous voyons que le discours des taxonomistes vis-à-vis des autres disciplines est ambivalent : elles sont tour à tour adversaire, bénéficiaire ou fournisseur de la taxonomie. Ce travail de thèse, et plus particulièrement les cartographies effectuées dans les chapitres 5 et 6, a mis en lumière que le domaine précédemment défini comme étant la taxonomie – le travail sur des collections à but nomenclatural – a peu de liens avec les autres spécialités. Cette séparation est d'autant plus forte que le savoir taxonomique d'un groupe est faible : la délimitation est en effet beaucoup plus nette pour la carte 5.6 issue des travaux sur les hémiptères que pour la carte 5.2 issue des travaux sur les oiseaux. Ce n'est donc pas (encore ?) une discipline de synthèse, mais une discipline dont le travail est situé en amont de toute autre étude.

Nous avons vu partie III qu'il n'existait pas de déclin global de la discipline mais qu'il semblerait en revanche s'opérer une redistribution des forces taxonomiques. Nous avons également vu partie IV que l'inquiétude au sujet de la discipline revenait de façon récurrente dans la littérature. Faut-il conclure pour autant que ces inquiétudes n'ont pas de raison d'être ?

Cinquième partie

Faut-il développer la taxonomie ?

Chapitre 12

Importance de la taxonomie

12.1 Pourquoi développer la taxonomie ?

Il existe deux sortes d'arguments en faveur de la taxonomie :

arguments pragmatiques : elle est indispensable, d'une part, aux autres disciplines biologiques ; et, d'autre part, à la société ;

arguments ontologiques : elle contribue à faire avancer la connaissance humaine ; que ces connaissances soient reprises ou non par la suite.

Parmi les arguments pragmatiques, on trouve : son importance pour la biologie (Robinson, 1901; Anonymous, 1946b; Holyński, 2008) et en particulier pour l'étude des relations évolutives (Anonymous, 2002, p.ex.), pour l'écologie (Cockerell, 1946; Graham, 1972; Tautz *et al.*, 2002, p.ex.) et pour la conservation (Kosztarab, 1984; Dubois, 2003b, p.ex.). On trouve également des arguments soulignant son importance pour la recherche médicale et la santé (Pallini *et al.*, 2007, p.ex.), ainsi que pour la domestication (Shantz, 1934, p.ex.) ou encore pour l'économie, étant donné l'importance de la biodiversité pour cette dernière (Costello *et al.*, 2006, p.ex.).

Les auteurs défendant la taxonomie en tant que science dont les questionnements internes sont suffisants pour justifier de son existence, sont également nombreux. Citons, par exemple, Stirton *et al.* (1990); Bremer *et al.* (1990).

Un troisième type d'arguments, que l'on retrouve moins fréquemment, est celui de l'épanouissement personnel et sociétal procuré par l'étude de la nature. Un exemple parmi les plus clairs de ce discours est extrait d'un ouvrage du paléontologue Georges Cuvier (1769-1832) :

Je crois qu'elle [la connaissance de l'histoire naturelle] peut aussi contribuer à l'adoucissement des mœurs et au bonheur individuel. Ceux qui s'occupent paisiblement de l'étude de la nature, doivent être peu tentés de

se lancer dans la carrière orageuse de l'ambition; ils doivent succomber difficilement aux passions brutales ou cruelles, écueils ordinaires de ces têtes ardentes, qui ne savent pas maîtriser leur activité; purs comme les objets de leurs recherches, ils doivent être animés, pour tout ce qui les entoure, de cette même bienfaisance qu'ils voient exercer par la nature envers toutes ses productions. (Cuvier, 1798, p. x-xii)

Cet argument a été étudié de manière philosophique et je ne le développerai pas ici. On peut le rencontrer aujourd'hui, sous une forme quelque peu différente : « *Society has a growing need for credible taxonomic information in order to allow us to conserve, manage, understand, and enjoy the natural world* » (Wheeler *et al.*, 2004).

La dichotomie entre argumentation pragmatique et argumentation ontologique découle directement de la double nature de la taxonomie : une discipline productrice de savoirs obtenus pour sa propre curiosité et répondant également à des besoins qui lui sont étrangers.

Sur l'ensemble de la littérature que j'ai pu traiter, j'ai constaté que l'argument le plus fréquemment employé était celui de l'utilité de la taxonomie pour les autres sciences. La biologie de la conservation vient largement en tête, suivie de l'écologie. Plusieurs travaux insistent sur l'importance de la taxonomie pour la conservation à l'aide d'exemples concrets : par exemple Townsend (2006); Spartaco et Giovanni (2007); Paquin *et al.* (2008); Bouchet *et al.* (2009).

Cet argument est très répandu mais il n'existe pas d'étude ayant évalué concrètement cet apport de la taxonomie aux autres disciplines. Étant donné la couverture estimée de la taxonomie par le *Web of Science* (dans le meilleur des cas, environ 52 %, cf. section [Estimation de la couverture taxonomique du WoS](#) page 74), il semble délicat d'évaluer quantitativement cet apport à travers les citations que les articles font et reçoivent. Si l'utilisation du *Web of Science* était possible dans un cadre de cartographie qualitative, il ne semble pas recommandé de l'utiliser dans un cadre strictement quantitatif. Un autre moyen d'évaluer l'apport de la taxonomie aux autres disciplines consisterait à étudier l'emploi que ces disciplines font des noms d'espèces - ou des concepts d'espèces - créés. C'est ce que nous allons voir dans la section suivante.

12.2 Mesure de l'apport de la taxonomie aux autres disciplines

Dans la partie suivante de ce travail, j'ai évalué la proportion d'espèces différentes étudiées dans des disciplines non taxonomiques afin de répondre à deux questions :

12.2 Mesure de l'apport de la taxonomie aux autres disciplines

(1) ces espèces sont-elles nombreuses ? (2) Sont-elles toujours les mêmes, quelle que soit l'époque ?

12.2.1 Méthodologie

J'ai téléchargé les références sous forme tabulée de tous les articles indexés dans le *Zoological Record* en 2008¹, c'est-à-dire 76 567 références. Après les avoir importées dans R et ôté les doublons, leur nombre se réduisait à 76 486. J'ai considéré comme doublons les références pour lesquelles les auteurs, l'année, le titre et la source sont identiques (à l'exclusion des références dont le titre est inconnu : *[Title unknown.]*).

J'ai ensuite travaillé sur le champ DE (Descriptors Data), dont un exemple est rappelé ci-après, afin d'extraire tous les taxons mentionnés dans chaque document par l'utilisation successive d'expressions régulières appropriées.

Pseudodiplorchis americanus [Amphibian hosts / / Scaphiopus couchii ; Host mating success reaction ; test of sexual selection hypothesis] [Prevalence / / Scaphiopus (Amphibia) ; host mating success relationship] [Reaction of host to parasite / / Scaphiopus (Amphibia) mating success ; test of hypothesis] [Arizona / / Amphibian host mating success reaction ; field & laboratory study]. Scaphiopus couchii [Mating / / Male success ; effects of monogenean parasite burden] [Pair formation / / Female mate choice ; male monogenean parasite burden relationship] [Breeding success / / Males ; monogenean parasite burden relationship] [Natural selection / / Parasite mediated sexual selection ; test of hypothesis] [Monogenean parasites / / Pseudodiplorchis americanus ; Male mating success relation ; rigorous test of sexual selection hypothesis] [Prevalence / / Pseudodiplorchis (Monogenea) male mating success relationship] [Reaction of host to parasite / / Male mating success ; Pseudodiplorchis (Monogenea)] [Arizona / / Monogenean parasite effects on mating success ; field & laboratory study].

J'ai enfin listé tous ces noms et calculé leur fréquence d'utilisation. Les noms d'espèces ou sous-espèces ont été identifiés en repérant par expressions régulières les formes typographiques suivantes :

Acentrogobius simplex
Acanthogyrus (Acanthosentis) tilapiae
Abyssochrysos ? giganteum
Salmo trutta labrax
Agaricia agaricites form. masiva
Theronia (Theronia) clathrata flavolineata

1. requête « IC = "ZOOR*" AND PY = 2008 ».

Anopheles gambiae species complex

J'ai ensuite réalisé l'opération pour deux années supplémentaires, choisies en fonction de leur place dans la distribution du *ZR* : 1980 et 1998 (voir la section [Distribution des articles dans le *ZR* page 28](#)). Notons que ces milliers de noms extraits du *Zoological Record* n'ont pas été vérifiés un à un.

12.2.2 Résultats

Le champ DE est vide pour 1 % des références. Lorsqu'il est complété, il y est toujours fait mention d'au moins un taxon. La moyenne du nombre de taxons mentionnés est de 1,77. Le nombre de publications indexées comme étant des articles de systématique est de 10 575. Le nombre de publications ne comportant pas le terme *Systematics* dans le champ BD est 65 911.

Cas des articles de systématique

Dans les publications de systématique, 31 360 noms de taxons sont mentionnés dans le champ DE. Parmi eux, 23 859 sont des noms d'espèces ou sous-espèces, répartis dans 7 822 documents (soit 74 % des documents).

Cas des autres articles

Dans les publications différentes de celles de systématique, 31 612 noms de taxons distincts sont mentionnés dans le champ DE. Parmi eux, 26 683 sont des noms d'espèces ou sous-espèces, répartis dans 42 509 articles (soit 55 % des documents). Sur ces 26 683 noms, 2 720 correspondent à des études en biologie de la conservation. Si l'on se rapporte aux données de l'Index to Organism Names, 1 130 569 espèces ont été décrites jusqu'en 2007 inclus. Ainsi, en 2008, 2,4 % de ces noms étaient utilisés et 2 ‰ traités en biologie de la conservation.

Un faible pourcentage de noms d'espèces et sous-espèces (4,5 %) sont mentionnés à la fois en systématique et dans d'autres disciplines. Si l'on regarde de plus près, on constate que 45 % d'entre eux font en fait référence à des taxons hôtes : ce sont donc leurs parasites qui sont étudiés de manière taxonomique. Par conséquent, moins de 2,5 % des espèces et sous-espèces sont étudiées en systématique et dans une autre discipline. Elles appartiennent autant aux vertébrés qu'aux invertébrés : Insecta (25 %), Mammalia (18 %), Pisces (13 %), Arachnida (7 %), Crustacea (7 %), Aves (7 %), Reptilia (6 %), Mollusca (5 %), Amphibia (4 %).

12.2 Mesure de l'apport de la taxonomie aux autres disciplines

Le tableau 12.1 présente la liste des 20 noms d'espèces ou sous-espèces les plus utilisés, ainsi que leurs occurrences. Sans surprise, on constate que ce sont des espèces d'intérêt pour l'homme : organismes modèles, vecteurs de maladie, ravageurs, espèces d'élevage ou emblématiques (renard, éléphant, cerf...). On peut être surpris de constater que *Drosophila melanogaster* ne soit mentionnée que dans 223 cas sur l'ensemble des 76 486 documents. On peut supposer qu'il est fréquent que le nom latin ne soit pas mentionné dans le cas d'organismes modèles.

NOM LATIN	NOM VERNACULAIRE	OCCURRENCES
<i>Drosophila melanogaster</i>	mouche du vinaigre	223
<i>Apis mellifera</i>	abeille européenne	177
<i>Plasmodium falciparum</i>	-	161
<i>Salmo salar</i>	saumon atlantique	154
<i>Gadus morhua</i>	morue de l'atlantique	145
<i>Aedes aegypti</i>	-	142
<i>Oncorhynchus mykiss</i>	truite arc-en-ciel	133
<i>Oreochromis niloticus</i>	tilapia du Nil	121
<i>Mytilus galloprovincialis</i>	moule méditerranéenne	120
<i>Vulpes vulpes</i>	renard roux	120
<i>Pan troglodytes</i>	chimpanzé	115
<i>Cervus elaphus</i>	cerf élaphe	114
<i>Sus scrofa</i>	sanglier	113
<i>Loxodonta africana</i>	éléphant de savane d'Afrique	112
<i>Toxoplasma gondii</i>	-	111
<i>Canis lupus</i>	loup	110
<i>Odocoileus virginianus</i>	cerf de Virginie	110
<i>Helicoverpa armigera</i>	noctuelle de la tomate	109
<i>Salmo trutta</i>	truite	109
<i>Anopheles gambiae</i>	moule commune	106

TABLEAU 12.1 – Liste des 20 espèces les plus fréquemment étudiées dans des études non taxonomiques en 2008.

La figure 12.1 présente la fréquence d'utilisation des 26 683 noms d'espèces et sous-espèces distincts présents dans les documents non systématiciens du *Zoological Record* en 2008. On observe qu'une minorité de noms est utilisée fréquemment, et que la majorité n'est utilisée que très peu de fois.

Ainsi, ce n'est pas une minorité de noms qui est utilisée de manière intensive en biologie, mais une majorité de noms qui est utilisée de manière ponctuelle.

Si l'on effectue les mêmes analyses en 1980 et 1998, on obtient les mêmes courbes de distribution (résultat non montré) et le même ordre de grandeur pour la proportion de noms décrits utilisés : 2,8 % pour 1980 et 2,5 % pour 1998.

Parmi tous ces taxons étudiés en 2008, 4 020 l'étaient également en 1980 et en 1998. Cela représente 28 % des taxons étudiés en 1980 et 15 % des taxons de 2008.

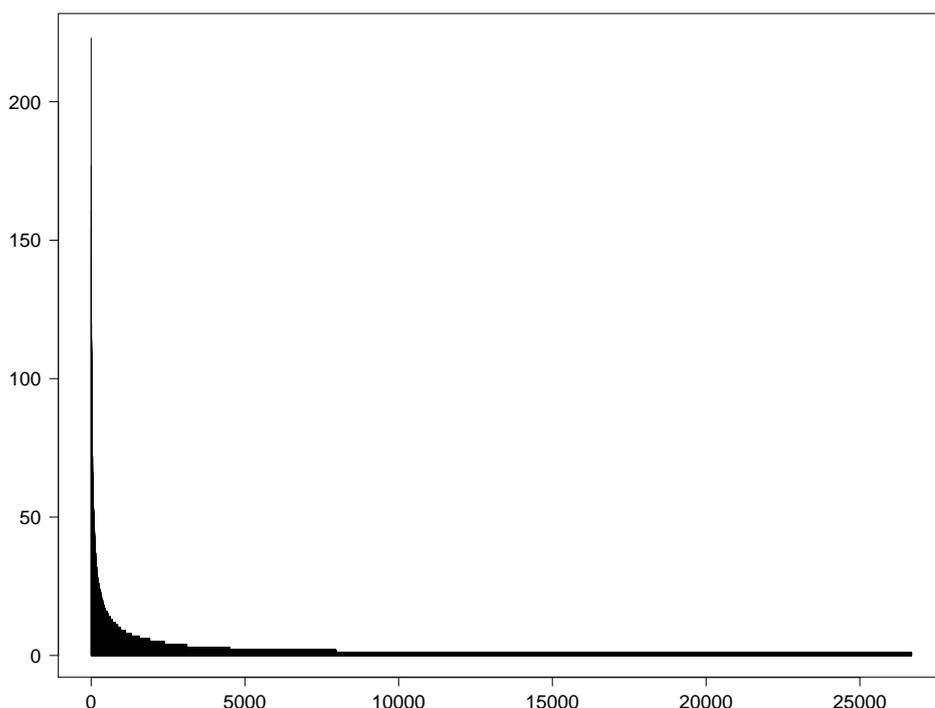


FIGURE 12.1 – *Fréquence d'utilisation de 26 683 noms d'espèces et sous-espèces distincts dans les documents non systématiciens du Zoological Record en 2008. Source : champ DE du Zoological Record.*

12.2.3 Conclusion

Sans choix d'une mesure étalon, il est difficile de savoir si ce pourcentage d'utilisation de 3 % des noms dans des études non taxonomiques est un chiffre élevé ou non. Ce pourcentage global paraît faible à première vue, mais il ne faut pas oublier qu'il fait référence à des milliers d'espèces et sous-espèces. En revanche, du fait que ce ne sont pas toujours les mêmes taxons qui sont étudiés d'une année sur l'autre, et que ce taux de taxons communs est faible ($< 30\%$), on peut considérer que la taxonomie joue un rôle important, et un rôle qui n'a pas perdu de sa vitalité : donner aux autres sciences le vocabulaire nécessaire pour désigner leurs objets d'étude.

Toutefois, les chiffres obtenus en ce qui concerne plus particulièrement la biologie de la conservation (2 ‰ des espèces et sous-espèces étudiées), ajoutés à ceux obtenus en ce qui concerne l'interdisciplinarité (5,4 % de tous les travaux pluridisciplinaires de la Systématique sont réalisés avec la biologie de la conservation, voir page 92) et enfin, à ceux de l'UICN (environ 60 000 espèces dont le statut de conservation a été évalué²), remettent en cause la justification de la taxonomie par son apport à la biologie de la conservation.

2. <http://www.iucnredlist.org/about/red-list-overview>, section *Expanding the Taxonomic Coverage of the Red List*.

Chapitre 13

Proposition de mesures de sa santé

13.1 Quand *inertie* remplace *déclin*

Replaçons maintenant la question de la mesure de la vitalité de la taxonomie dans son ensemble. C'est une question qui présuppose (1) qu'il existe un état de la discipline que l'on peut qualifier de « bon » et un état que l'on peut qualifier de « mauvais » ; et que (2) ces états peuvent être accessibles à travers des mesures quantitatives. Or, que serait un « bon état » de la taxonomie ? Il n'en existe pas de définition absolue : cette définition varie en fonction de la personne qui s'interroge et du critère qu'elle choisit d'utiliser. On pourrait considérer, par exemple, que la taxonomie est en bonne santé si :

- elle produit autant ou plus de connaissances que ce qu'elle a pu produire à d'autres époques ;
- elle produit autant ou plus de connaissances au regard des objectifs qu'elle se fixe ;
- elle produit autant ou plus de connaissances au regard des objectifs que ses utilisateurs lui fixent ;
- elle produit autant ou plus de connaissances au regard des investissements reçus ;
- elle reçoit suffisamment d'investissement au regard des objectifs qu'elle se fixe et de l'importance qu'elle a ;
- elle reçoit suffisamment d'investissement au regard des objectifs et de l'importance que ses utilisateurs lui donnent ;
- elle croît autant que les autres disciplines ;
- elle innove régulièrement et connaît des avancées conceptuelles et méthodologiques importantes ;
- ...

Si l'on résume, le seuil limite qui départagerait les deux états pourrait être défini en fonction (1) d'un **état antérieur** (ex. : plus ou moins d'espèces décrites?); (2) d'un **objectif**, fixé de manière **interne** ou **externe** à la discipline (ex. : décrire toutes les espèces avant qu'elles disparaissent); (3) d'un **critère de rentabilité** (ex. : nombre d'espèces décrites en fonction du nombre de taxonomistes) ou (4) des **autres disciplines**. Les mesures, quant à elles, porteraient sur (1) les **productions** de la taxonomie ou (2) les **investissements** qui y sont faits.

L'approche précédente a consisté à effectuer des mesures sur ses productions et indirectement, sur ses investissements; en prenant comme points de référence : un état antérieur de la discipline, d'une part; et les disciplines voisines, d'autre part.

Or, nous avons vu que la taxonomie est avant tout définie comme une discipline au service de la science et de la société. Par conséquent, le point de référence qui revient le plus fréquemment est celui de l'objectif qu'elle se fixe : **décrire l'ensemble de la biodiversité**. Régulièrement, des interrogations ou des certitudes se posent quant à son avancée :

What is the present strength of this all-important foundation [classification] ? Is it built upon rational principles ? Should we build on or tear down and reconstruct ? Is it nearing completion or does it represent as yet only the' earliest stages of the desired structure ? (Robinson, 1901)

"alpha taxonomy" is far from completed (Steyskal, 1965)

most of the terrestrial flora and the terrestrial and freshwater invertebrates have not been described taxonomically (Kosztarab, 1984)

most of the world species still to be documented (Jarzembowski, 1990)

Ces interrogations sont rendues plus aigües avec la prise de conscience que des espèces disparaissent :

Probably the highest-priority task facing systematic biology in general is completion of the inventory of the world's organisms before so many are destroyed that a reasonable picture of them can no longer be obtained (Thimann et al., 1966)

the classification and description of the approximately 200,000 species of flowering plants is still far from complete. Today, the need for this is ever more urgent, as pressure on natural environments becomes more acute (Harley, 1990)

In view of the available taxonomic manpower and the enormous rate of extinction that will characterize the next century, it is doubtful that even

13.1 Quand inertie remplace déclin

5 percent more of the world's organisms can be added to our inventory before the remaining 80 percent becomes extinct (Raven, 1972)

extinctions are occurring before even half of the world's species have been described (Costello et al., 2006)

Aujourd'hui, le point de référence est donc celui d'un objectif particulier : **décrire l'ensemble de la biodiversité actuelle avant que celle-ci ne soit définitivement inaccessible.**

C'est ainsi qu'avec les notions de *crise de la biodiversité* (*biological diversity crisis*, Wilson, 1985a), *siècle des extinctions* (*century of extinctions*, Dubois, 2003b), apparaissent celles d'*urgence taxinomique* (*taxonomic urgency*, Dubois, 2010) ou du besoin de révolutionner la taxonomie (Godfray, 2002c). L'emploi d'un registre lexical dramatique fort, dénonciateur de la situation de la taxonomie serait donc dicté non pas par l'ampleur effective de l'extinction de la taxonomie, mais par l'ampleur estimée du *fossé taxinomique* (*taxonomic gap*, Dubois, 2010), i.e. « *the wide discrepancy between the reality of specific biodiversity and our knowledge of it* ». Ce serait l'ampleur supposée de la crise de la biodiversité, et les inquiétudes qui en découlent, qui seraient à l'origine de cette thèse du « déclin » de la taxonomie.

Une des conséquences directes de ce choix de référentiel est que les investissements qui sont faits dans la discipline ne pourront jamais être considérés comme suffisants tant que ne sera pas établie ce que González-Oreja (2008) appelle l'*encyclopédie de la vie* (*Encyclopedia of Life*). Dans ces conditions, faut-il remplacer le terme *déclin* par celui d'*inertie* ? González-Oreja (2008) a modélisé le temps nécessaire pour parvenir à un inventaire complet de la biodiversité en considérant que nous connaissons aujourd'hui 1,5 millions d'espèces, décrites au rythme de 10 000 espèces par an. En faisant varier deux paramètres, l'estimation du nombre total d'espèces restant à décrire (entre 3 et 100 millions), et l'accroissement potentiel du nombre de taxonomistes (x2, x3, x5, x10, x15), il détermine qu'il faudrait des milliers de taxonomistes pour achever l'ouvrage en une génération : « *If the mean rate of description of new species is 5 per year per person (indeed, an optimist situation, but optimistically not impossible to achieve), then some 66,000 taxonomists are required; if the mean rate of description is 2.5 (the highest current value), some 132,000; and if it is only 1.4 (the lowest current value), some 236,000 taxonomists are desperately needed, right now, and every year during the following 25 years* » .

Bien que les données du chapitre précédent ne permettent pas de supporter de manière quantitative l'argument pragmatique de l'utilité de la taxonomie pour les autres disciplines, et plus particulièrement pour la biologie de la conservation, elles n'invalident pas pour autant l'argument ontologique en faveur du développement de la taxonomie dans ce nouveau paradigme d'extinction des espèces auquel la biologie

fait face (Dubois, 2008b). L'évolution étant un processus historique, imprévisible et unique, le seul moyen d'avoir accès à la connaissance de la biodiversité est de l'échantillonner. Par conséquent, *si l'on souhaite poursuivre notre exploration de la vie sur cette planète*, il est indispensable de promouvoir le développement de la taxonomie aujourd'hui.

S'il est hors de propos de ce travail de trancher en faveur de l'une ou l'autre des solutions proposées pour tonifier la taxonomie (voir références en introduction), il permet toutefois de proposer de nouvelles mesures pour évaluer la vitalité de cette discipline.

13.2 Évaluer la taxonomie à travers ses collections

En tenant compte de la définition donnée précédemment à la taxonomie – l'étude de collections naturalistes dans un but nomenclatural – je propose de mettre en place un suivi de la santé de la discipline à travers des mesures portant sur ses collections. L'informatisation croissante de celles-ci, l'existence de structures fédératrices telles que la NSCA (*Natural Science Collections Alliance*) ou le GBIF (*Global Biodiversity Information Facility*) sont des outils privilégiés pour centraliser des informations de cette nature. L'importance des collections a récemment été défendue à plusieurs reprises (voir par exemple Suarez et Tsutsui, 2004; Szwedo et Iwan, 2008; Hoberg *et al.*, 2009).

On pourrait ainsi envisager d'établir des mesures sur :

les spécimens : quel est le rythme d'accroissement des collections ? Quel est le nombre de spécimens n'ayant jamais été étudiés ? Étudiés une fois ? Deux fois ? Apparaissant dans une publication ?

les ressources humaines : de combien de spécimens doit s'occuper un technicien ?

les visiteurs : la collection reçoit-elle souvent des visiteurs ?

Ces propositions ne sont qu'une ébauche et nécessiteraient d'être développées avec des acteurs en charge des collections. Contrairement aux estimations basées sur le nombre d'espèces décrites et/ou restant à décrire, ces mesures établiraient un lien étroit entre les données disponibles et l'utilisation qui en est faite. Ceci apparaît comme le meilleur moyen d'avoir un aperçu clair, concret et direct des besoins d'échantillonnage et de révision taxonomique. Cela pourrait ainsi servir de guide à l'élaboration de stratégies de collecte et d'étude.

Pour finir, il n'existe à l'heure actuelle aucune base de données fiable permettant d'estimer précisément le nombre de taxonomistes dans le monde. L'histoire des annuaires de systématiciens est assez longue (citons les deux plus importants : *The*

13.2 Évaluer la taxonomie à travers ses collections

naturalists' directory, qui paraîtra irrégulièrement, totalisant 46 éditions (a priori jusqu'en 1993) ; ou encore le *Directory of Zoological Taxonomists of the World* de 1959). Cependant il n'y a jamais eu d'initiative continue jusqu'à aujourd'hui. Depuis 1990 de nombreuses bases de données informatisées ont vu le jour. La seule base à contenu et vocation réellement internationaux est la *World taxonomists database* (n > 4 800) qui est fondée en 1995 par ETI BioInformatics, une organisation non gouvernementale en lien avec l'UNESCO. Elle est alimentée par l'inscription des taxonomistes eux-mêmes. Toujours dans une perspective d'évaluation de la santé de la discipline, il pourrait être demandé à toute personne travaillant sur des collections dans un but nomenclatural de s'inscrire sur cette base. Il pourrait alors être fait, pour la première fois, un inventaire exhaustif des taxonomistes et de leurs compétences respectives, afin d'analyser la répartition des forces en présence aujourd'hui.

Conclusion

Ce travail est le premier à analyser finement la taxonomie en se basant sur le *Zoological Record*. Il a donné lieu à la mise en place d'outils d'analyse et de traitement des données sous **R** : import des données, déduplication, extraction de mots-clés, gestion des duplicatas d'auteurs.

Au cours de ce travail nous avons vu que la taxonomie était une discipline définie de deux manières. La première présente la taxonomie comme un prestataire de services (réaliser un inventaire, classer, donner des noms, créer un système prédictif, rendre accessible l'information) tandis que la seconde la présente comme une science à part entière, avec ses propres questionnements et paradigmes (refléter les relations évolutives, comprendre la flore et la faune, explorer la planète, théoriser la classification, etc.). Ces différentes définitions peuvent être interprétées comme une marque de son dynamisme. La réalisation de trois cartographies de la zoologie basées sur la cooccurrence de mots-clés m'a permis de proposer une nouvelle définition de cette discipline, comme *branche de la biologie travaillant sur des collections dans un but nomenclatural*. Cette définition peut sembler réductrice, reflet d'un choix de publication mais non reflet d'une réalité de travail. J'ai donc vérifié qu'elle se retrouvait également au niveau des auteurs.

Des mesures effectuées sur le *Zoological Record* ont montré que cette discipline n'est pas en déclin, contrairement à ce qui est affirmé de manière répétée dans la littérature. Il semblerait qu'on assiste aujourd'hui à une redistribution des forces de la discipline, notamment en ce qui concerne les pays impliqués dans ces recherches, avec une part importante prise par les pays émergents.

Il s'est en outre avéré que ce discours autour du déclin est ancien puisqu'on le rencontre déjà au début du XX^e siècle, et de manière régulière tout au long du XX^e siècle. Il s'accompagne d'arguments récurrents mettant en opposition la taxonomie et les autres disciplines, dont l'apparition est tantôt perçue comme un obstacle, tantôt perçue comme une opportunité d'accéder à de nouvelles données. Il semblerait toutefois que dans la pratique, la taxonomie soit relativement isolée de ces nouveaux champs, sauf pour des taxons comme celui des Oiseaux, dans lesquels elle est bien connue.

La prévalence de ce discours et la redondance des arguments qui l'accompagnent ne permettent toutefois pas de relativiser l'importance de développer la taxonomie aujourd'hui. Il faut en effet prendre un nouveau paramètre en compte, le paradigme de l'extinction des espèces. Des mesures sur l'utilité de la taxonomie pour la biologie de la conservation montrent que nous sommes loin d'avoir une synergie entre les deux disciplines. S'il paraît difficile, voire impossible, de pallier la disparition des espèces, il est peut-être encore temps de diminuer la menace qui pèse sur la taxonomie. En effet, outre la biodiversité, c'est également une science qui est menacée à travers la disparition de ses objets d'étude.

Dans ce contexte, je propose d'évaluer la vitalité de la taxonomie non pas en fonction de ses publications, mais plutôt en fonction de la vitalité de ses collections. La mise en place d'indicateurs estimant leur accroissement et leur degré d'utilisation apporterait un éclairage très instructif sur sa santé.

Perspectives de recherche

Ce travail ouvre plusieurs perspectives de recherches, dans différents champs disciplinaires. Ce questionnement sur l'état de la taxonomie est en effet fondamental puisqu'il se situe à la croisée des chemins entre science, société, philosophie et politique.

Il serait tout d'abord intéressant d'étudier de manière plus approfondie les discours sur le déclin de la taxonomie dans une perspective historique. Une analyse des discours introductifs des congrès annuels de plusieurs sociétés savantes serait certainement très instructive.

Il serait également pertinent d'étudier par une approche sociologique ou anthropologique la dynamique de la communauté des taxonomistes au sein d'un établissement scientifique, en la comparant avec celle d'un autre établissement ou encore avec celle d'une autre discipline. Ce sujet de thèse ayant pris naissance au Muséum national d'Histoire naturelle de Paris, il serait d'autant plus intéressant que cet établissement en fasse partie. Cela permettrait de faire pour la première fois la relation entre les différents acteurs mis en cause dans le discours sur le déclin : étudiants, taxonomistes, autres scientifiques, politiques, utilisateurs finaux, chargés de collection. Cela serait également l'occasion de confirmer – ou infirmer – les arguments de type psychologisant tenus par de nombreux auteurs sur les taxonomistes, les dépeignant comme des personnalités individualistes, incapables de se réunir autour d'un objectif commun.

Un autre point important concerne le débat portant sur la technicisation de la science, et son impact sur la production de connaissances taxonomiques. Si cette thématique a été traitée de manière importante en philosophie des sciences, il serait

Conclusion

intéressant de cerner par une approche scientométrique l'influence réelle de nouvelles techniques sur la découverte et la description de nouvelles espèces.

Annexe A

Article de *Frontiers in Zoology*



DEBATE

Open Access

Insights from *Zootaxa* on potential trends in zoological taxonomic activity

Elise Tancoigne^{1*}, Cyprien Bole¹, Anne Sigogneau² and Alain Dubois¹

Abstract

Background: An opinion currently shared by taxonomists and non taxonomists alike is that the work of inventorying biodiversity is unbalanced: firstly, in favour of countries in which taxonomy has been studied for a long time, and, secondly, in favour of vertebrates. In the current context of threats of species extinction, access for taxonomists to biological material and information becomes crucial if the scientific community really aims at a better knowledge of biological diversity before it is severely and irreversibly impoverished. We performed an analysis of 748 papers published in *Zootaxa* in 2006 and 2007, as well as 434 questionnaires sent to their authors to test these opinions. A generalization of these results to zoological taxonomy as a whole is discussed.

Discussion: We found that the disequilibrium is not exactly what it usually considered to be. The USA, China and Brazil are currently the three leading countries in zoological taxonomy. Each of them presents, however, a different pattern. Taxonomists from Asia and South America are younger and mainly work in universities, not museums. A bias in favour of vertebrates still exists if we refer to the effort invested in each group to produce taxonomic data, but not to the number of papers. Finally, we insist on the idea that "describing a species" is very different from "knowing a species".

Summary: The *taxonomic involvement* of a country, in terms of manpower and funding, appears to be a key factor in the development of fruitful taxonomic research. This message seems to have been understood by the countries that recently decided to increase considerably their taxonomic involvement. It still has to be received by those who did not.

Background

Taxonomy is the disciplinary field of biology in charge of collecting, studying, describing, classifying and naming living organisms and taxa. It is a fundamental discipline for biology as a whole. It is thus crucial that taxonomists be given access to biological material and information. Previous studies [1-4] and widespread opinions [5,6] nevertheless postulate that access to specimens is unbalanced, both as regards the different regions of the world and the taxonomic groups. First, countries and areas where researchers have historically worked for a longer time on zoological taxonomy (Europe, North America and Australia) would therefore dominate zoological taxonomy. Second, the attention

paid to Vertebrata would be higher than for other taxa, especially Arthropoda.

Yet, a quick search of all papers describing new species in the *Web of Science* in 2006 reveals that Brazil and China are among the top five publishing countries (USA: 25%, China: 13%, Germany: 8%, Japan: 7%, Brazil: 6%).

So, as concerns the distribution of zoological taxonomy, where are we now? No recent quantitative study exists at the world scale. Such surveys were either limited to a taxon [3,7-11], or a country [12-22], or they offered a summary of fragmented information [23-25].

However, today a mega-journal of taxonomy, *Zootaxa*, provides a worldwide coverage and deals with numerous zoological taxa. In 2004, it published far more pages than the combined total published in the ten core journals in systematic zoology [26]. It contributed 13% of all new taxa of animals indexed in the *Zoological Record* since 2004 [27]. An analysis of this journal could

* Correspondence: tancoigne@mnhn.fr

¹UMR 7205 OSEB, Département Systématique et Evolution, Muséum national d'Histoire naturelle, CP 30, 25 rue Cuvier, 75231 Paris Cedex 05, France
Full list of author information is available at the end of the article

therefore produce a useful and relevant insight on today's taxonomical practices in zoology.

Quantitative tools will be used in this work to answer two questions:

1. Is zoological taxonomy still mostly practised in countries that have historically been taxonomically active?
2. Is zoological taxonomy still unbalanced in favour of vertebrates, *vs.* arthropods?

In conclusion, we will discuss the possible generalization of our results to taxonomy as a whole.

We determined the countries and areas in which (1) more works are published in zoological taxonomy; (2) most of the active taxonomists are working; (3) foreign taxonomists are mainly received; (4) taxonomists are involved in collaborations; (5) more species are described from other areas; (6) types specimens are kept; and, finally, (7) research is carried out in museums, which are traditional institutions for taxonomic research [9,15,28-30].

The disequilibrium in favour of vertebrates can be documented in terms of either taxonomic production (publications, pages, new species), or in the means employed for the studies (number of taxonomists, of specimens, of characters...). Therefore, we gathered additional information on (8) the taxa to which are devoted the higher number of publications, pages, new species as well as the taxa represented by the highest numbers of (9) taxonomists, (10) characters and (11) specimens used.

Methods

Sampling procedure

We chose to work on the years 2006 and 2007 because they were those when most articles had been produced at the time of our study ($n = 2138$). We randomly selected 748 articles among all those published in *Zootaxa* during this period. We devised a method to fix the sampling size. This method is explained below. Then we used three main procedures to obtain the data from our sample: data extraction from literature, data mining on databases and a questionnaire sent to each author of the sample who had provided an email. The data obtained from the articles or the questionnaire were entered into a database built using Access2002.

Five parameters describing each publication of the whole set of 2138 articles published in *Zootaxa* in 2006-2007 were available in the *Zoological Record online*: (1) treated taxon, (2) number of pages, (3) number of authors, (4) number of described species and (5) number of references. Three other parameters found on the *Zootaxa* website [31] were added: (6) whether the article is in open access or not, (7) its delay of publication after

submission and (8) its category (article or monograph). A sampling script was written with the R language [32]. We made 1000 random samplings of 5% of the total number of papers (function *sample()* without replacement of the default package *base*). This operation was repeated nine successive times with increasing sample sizes. Thus, we obtained 1000 samplings of 107 articles, 1000 samplings of 214 articles, etc. Then we used statistical tests to determine the rate of correct samples that each operation produced. Each of the 10,000 samples was tested and considered correct if the eight parameters did not significantly differ from the whole set ($P < 0.05$). As our data did not verify the normality condition, non-parametrical tests were performed for each of the eight parameters: χ^2 for (1); Wilcoxon/Mann-Whitney for (2), (3), (4), (5), (7); and binomial tests for (6) and (8). We decided to choose the minimal sample size for which 950 out of 1000 samples would not statistically differ from the whole set. This size is 35%, i.e., 748 articles (Figure 1). Finally, a random sampling of 748 articles was performed. We also checked that it did not significantly differ from the whole set for each of the eight parameters.

Questionnaire

The 748 articles studied were written by 1778 authors. Once multiple authorships were removed, 1384 authors remained, 1025 of whom had mentioned an e-mail

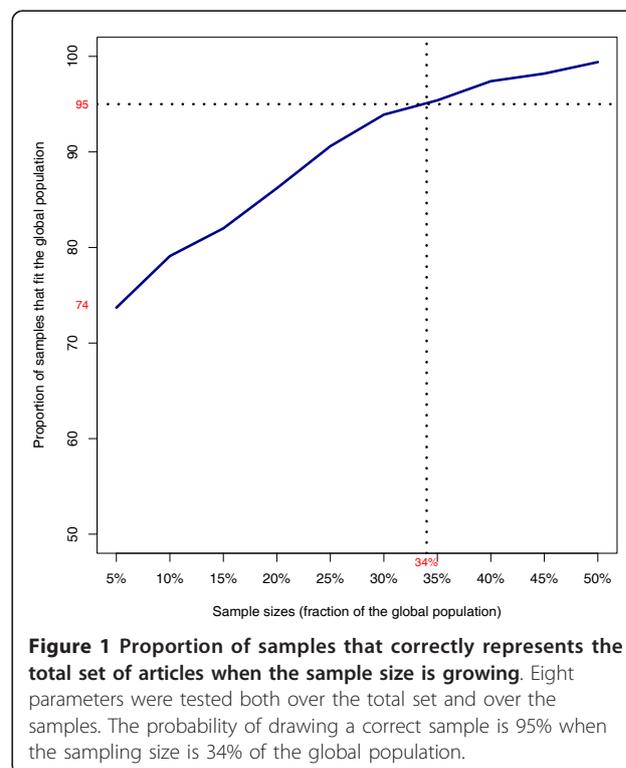


Figure 1 Proportion of samples that correctly represents the total set of articles when the sample size is growing. Eight parameters were tested both over the total set and over the samples. The probability of drawing a correct sample is 95% when the sampling size is 34% of the global population.

address in the paper. A questionnaire had been sent via e-mail in May 2008 to 10% of these 1025 authors for a preliminary study. It received 56 answers and comments from authors from most parts of the world (except Africa). Our questionnaire was considered approved and it was then sent once again in May 2009 to the 939 remaining authors. It was rejected by 105 mailboxes and received 434 answers in all (56 in 2008 and 378 in 2009). Its answer rate was thus 47%. We used Email-Questionnaire software to send our questionnaire [33].

Data picking

We obtained data for the 748 sampled articles; the 1778 authors and their 1111 institutions; the 1413 new species that were described in it and their 1404 holotypes; 15 syntypes; 23 468 paratypes and the 4987 other specimens used in descriptions. The 1778 taxonomic authors corresponded to 1384 “real persons”, of whom 434 taxonomists sent us personal data. Table 1 presents the data.

Each country was affiliated to a region through the data of the CIA world factbook [34]. Major regions are abbreviated the following way: Af (Africa), As (Asia), E (Europe), M (Middle East), N (North America), O (Oceania), S (South America). Given the small amount of papers from Central America, this region was grouped with South America.

Authors (i.e. names on paper) and taxonomists (i.e. real persons) were distinguished. For example, a taxonomist who writes two articles represents two authors.

Analyses

The number of publications per major region or country was computed fractionally. For example, an article written by two authors from Asia and one author who mentioned two addresses, one in South America and one in Africa, counts for 0.5 for Asia, 0.25 for South America and 0.25 for Africa. The number of publications per category of institution was also calculated fractionally. For example, an article written by two authors from the University of Sao Paulo and one author who mentioned two addresses, one in the University of Sao Paulo and one in the American National Museum, counts for 0.75 for Brazilian universities and 0.25 for North American museums. The number of type-specimens per category of institution is calculated directly. For example, a new species with one holotype kept in the University of Sao Paulo and 54 paratypes kept in the American National Museum counts 1 for North American museums and 54 for Brazilian universities. The number of taxonomists per major region or country was calculated fractionally once duplicated authors (authors with same name and address) had been removed. For example, if an author mentioned two addresses in two different major regions s/he counts 0.5 for each region. The collaboration

indicator of the *Observatoire des Sciences et des Techniques* (Paris, France) was used to study collaborations: (number of publications of a region A with an other region B)/(total number of publications of region A). For instance, an article written by one author from Asia and two from South America counts for one co-publication between Asia and South America.

All the data was analysed with the free statistical software R [32] and packages *ade4* [35], *RODBC* [36], *vcd* [37,38] and *FactoMineR* [39]. We used correspondence analysis (CA) to summarize the data's structure and identify similarities between the variables of a contingency table [40]. CA measures the distances between the row and column points and presents the inter-relations of variables in calculated dimensions that outlines maximum of variance (VARIMAX rotation). The analysis determines which category values lie close together and which are far from each other. These can be seen on a correspondence map, where the row and column categories are plotted along the computed factor axes. Spatial proximity between categories does not imply a correlation between the categories but instead a strong specific link in relation to these factors. This graphical method is based on a χ^2 test that measures the deviation to independence model. We also performed χ^2 statistical tests to compare proportions when it was needed.

Evaluation of *Zootaxa* biases

Our choice to work on *Zootaxa* inevitably represents a bias. Because our questions treat the different areas of the world and taxa, two possible biases concern the representativeness of (1) the taxa studied in *Zootaxa* and (2) the regions where the studies were carried out.

We compared the proportion of the different taxa studied in 2006 in *Zootaxa* and in all the publications indexed in the entire volume of the *Zoological Record* for 2006 (Figure 2). We observed that “taxo*”, “taxonomy” or “systematic” did not even allow us to find all articles published in *Zootaxa* in 2006. We used the keyword “sp nov”, coming from the *Zoological Record's* thesaurus, to find these publications in the *Zoological Record*. If all categories of Figure 2 are considered, a χ^2 test shows significant differences ($P < 0.001$) between the *Zoological Record* and *Zootaxa*. However, the distribution does not show significant differences ($P = 0.12$) if we consider only three classes of taxa: ‘Arthropods’, ‘Vertebrates’ and ‘Other’. Therefore, for the purposes of our study the taxa were grouped as ‘Vertebrates’, ‘Arthropods’ and ‘Other’.

We also compared the proportions of the different regions where new species were described in 2006 in *Zootaxa* and in all the publications with the same pattern indexed in the entire volume of the *Web of Science* (WoS) for 2006 (Figure 3). We used the *Analyze Results*

Table 1 Data extracted for each article, author, institution, new species and specimen involved

Article		
ZR ID number	Number attributed by the <i>Zoological Record</i>	ZR
Title		ZR
Abstract		ZR
Internet link	Internet link to the pdf article on the <i>Zootaxa</i> website	ZT
Number of pages	$n = (\text{ending page} - \text{beginning page}) + 1$	ZR + H
Delay of publication	Number of days between acceptance and publication	ZT + H
Category of document	<i>Monograph; article</i>	ZT
Category of work	We distinguished six categories of works. <i>additional data</i> : addition of sounds, karyotypes, redescrptions, chorology, phylogenies to the knowledge of a taxon, without new taxon description; <i>checklists or catalogue of species/types</i> ; <i>regional fauna</i> : taxonomic revision at regional scale; <i>taxonomic revision</i> : each taxon or types reviewed, or at least 3 different kinds of analyses done (e.g. morphological studies plus distribution plus karyotype of a taxon); <i>taxonomic work on isolated taxa</i> : single or multiple descriptions of new taxa, new descriptions plus new records of the taxon; <i>theoretical works</i> : biographies or nomenclatural changes only, or theoretical work on taxonomy and nomenclature	H
Authorship presentation	Whether authorship is in alphabetical order or not. Names beginning by "De", "Van", are considered at letters "D" or "V".	H
Phylum delimitation	Phylum as provided by ZR indexation	ZR
Taxon delimitation	The most inclusive taxon treated in the article	H
Level of work	Level of the most inclusive taxon treated in the article. <i>Species level</i> : any level up to and including species level; <i>genus level</i> : any level from species level up to and including genus level; <i>family level</i> : any level from genus level up to and including family level; <i>above family level</i> : any level above family level	H
Geographic delimitation	The most single inclusive area treated in the article. We used political boundaries of studied areas instead of biogeographical ones, in order to study the relationship between laboratories and studied areas	H
Level of geographic delimitation	We used four levels of area: <i>locality level</i> , <i>regional level</i> , <i>country level</i> , <i>major area level</i> . We used <i>other</i> when the study involves more than one major area. Oceans and seas were treated as regions and affiliated to a country or a continent when specified in the article	
Number of references		ZR
Price	Whether the article is in free access or not	ZT
Authors		
Family name		ZR
Firstname		ZR
Article related	List of the articles written by the author in <i>Zootaxa</i>	H
Position in the article		H
Institution	Institution of the author for the paper concerned. The ZR provides the address of the first co-author; the other ones were added from the paper itself. Whenever the author is affiliated to several institutions, this information was retained	ZR + H
Email	Some email addresses were provided by the ZR; others were found in the article	ZR + H
Professional status	<i>Amateur, Doctoral student, Established taxonomist researcher, Established non-taxonomist researcher, Master student, Postdoctoral, Retired researcher, Technician</i>	Q
Nationality	Country of birth	Q
Age		Q
Sex		Q
Highest degree	<i>Master (four or five years of faculty), PhD (or equivalent), Other, No degree</i>	Q
Free comment		Q
Institutions		
Name	The names of all institutions were checked to suppress duplicates	ZR + H
Type	Five categories of institutions were recognized: <i>Museum, University, Institute, Private, Other institution</i>	H
Country		ZR + H

Table 1 Data extracted for each article, author, institution, new species and specimen involved (Continued)

Continent	The affiliation of a country to a major area followed the CIA World Factbook data [34]	H
Level of the institution	Some institutions are subordinate to others, which may (or may not) appear in the address given by the author(s). For each of them, the following information was checked using their web sites: 1: whether the institution is autonomous 2: whether it is subordinate to another institution 3: whether it is subordinate to two other institutions	H
Subordinating institution	Name of the superordinate institution if relevant	H
New species		
Nomen		ZR
Specimens involved	List of the specimens involved	H
Allotypes	Whether allotypes are designated or not	H
Developmental stages	Whether numerous developmental stages are involved or not	H
Methods used for tree construction	<i>Parsimony, Bayesian/maximum likelihood, Distance, Combined methods</i>	H
Characters used	<i>Morpho-anatomical, Molecular, Bio-acoustic, Karyological, Other</i>	H
Specimens used to describe a new species		
Number of specimens	Number of specimens of each type for each new species	H
Kind of specimen	<i>Holotype, Syntype, Neotype, Paratype, Other specimen</i>	H
Place of collect	Collecting locality as stated in the paper, with GPS coordinates if available. Localities were distinguished through their names or through their GPS coordinates. Each locality was referred to a country and a realm whenever possible. The geographical delimitations used are the same as for the publications	H
Institution of conservation	Institution where specimen(s) is/are kept	H
Year of collect		H
Size (cm)	For holotypes only	H
Sex	For holotypes only. <i>Male; Female; Hermaphrodite; Unknown</i> when the information is lacking. Because many authors do not indicate the sex when they use juveniles, we also added the category <i>Juvenile</i>	H

H: data added by hand. Q: data from the questionnaire. ZR: data from the *Zoological Record* online. ZT: data from the *Zootaxa* website.

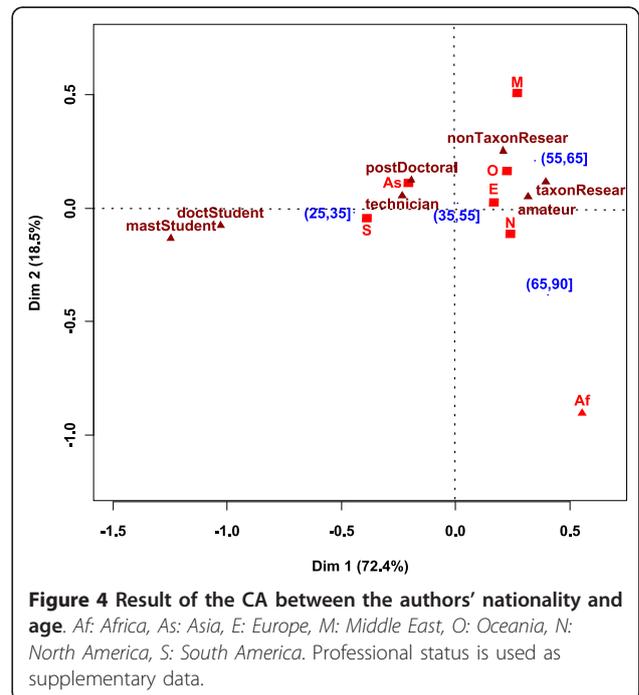
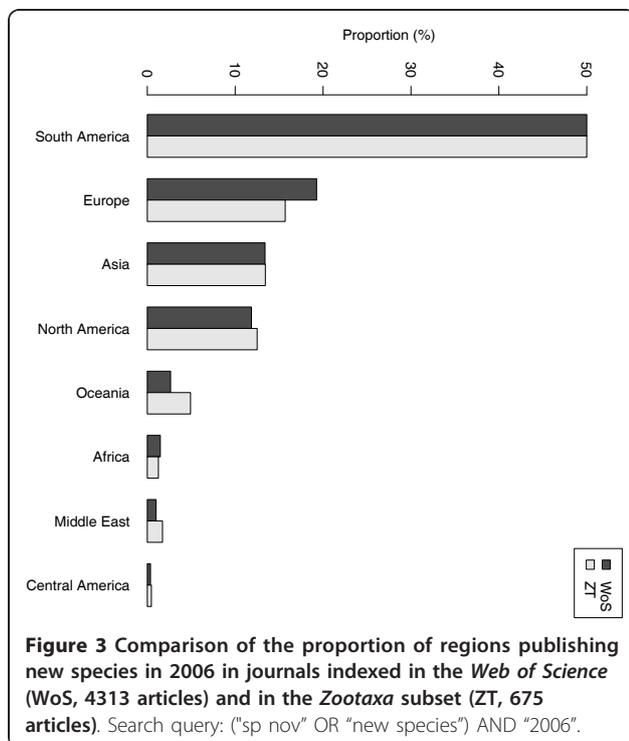
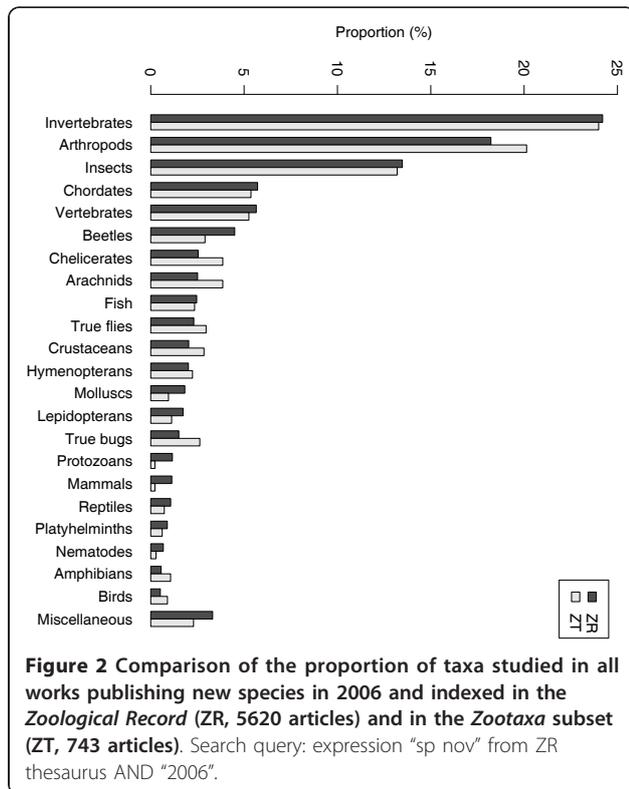
tool of the *Web of Science* which computes information on authors' countries. Our search query "sp nov" OR "new species" found 675 articles for *Zootaxa* and 4313 articles for the *Web of Science*. We did not use the truncated query "new sp" because it returned articles about language ("speech", "spoken"). If all the regions of Figure 3 are considered, a χ^2 test shows significant differences ($P < 0.001$) between the *Web of Science* and *Zootaxa*. However, the distribution does not show significant differences if we consider only three classes of regions: Asia-South America (AsS), Europe-North America-Oceania (ENO) and Africa-Middle East (AfM) ($P = 0.46$); or Africa-Europe-Oceania (AfEO), South America-North America (SN) and Asia-Middle East (AsM) ($P = 0.36$).

Results

Publications dealing with species are predominant in *Zootaxa* (74% of the works analysed). Only 17% of the papers deal with taxa between genus and family, and 9% with taxa above family level. Therefore, so-called " α -taxonomy", "microtaxonomy" [41], or "eidonomy" [42] is predominant. Works on isolated taxa (single species descriptions, etc.) are the most numerous (57%), followed by taxonomic revisions (17%), additions of new

data for taxa already known (11%), regional faunas (8%), checklists or catalogues of species/types (4%) and theoretical works (3%). These terms are defined in Table 1.

Most of the people who answered our questionnaire are fully qualified researchers (60%), either taxonomists (45% of total) or not (15%). Students come in second position, with 15% PhD students and 7% MSc students. Postdoctoral students account for 9% and retired researchers, amateurs and technicians for the remaining 9%. The proportions listed above vary between the three areas previously defined (Figure 4). Figure 4 shows the result of the CA between the age and the region of origin of the authors. Professional status was used as supplementary data. We deleted the 'retired researcher' category because it was naturally linked with the (65, 90] category and made the map confusing. Africa is distant on the map because it has few values (0, 2, 1, and 0) which produce high frequencies (0, 66, 33 and 0). Dimension 1 is strongly linked with the age and pits the (25, 35] category against the (55, 65] and (65, 90] categories. This is very interesting because this separation depends solely on the nationality categories. Indeed, (25, 35] is linked to S and appear to be the category with the highest proportion of S and the (55, 65] and (65, 90] categories are linked to M, O, N and appear to be the categories with



the highest proportion of them. Being at the center of the map, the (35, 55] category represents the average category as related to nationality. E represents 38% of (35, 55], S 23%, N 16%, As 15%, O 4%, M 2%, Af 1%.

Dimension 2 separates M from the other categories because no individuals belonging to the (65, 90] category answered our questionnaire. The supplementary data show that the (25, 35] category was related to technicians and postdocs, whereas (55, 65] was related to established taxonomic researchers, established non-taxonomic researchers and amateurs. These data suggest that the population of taxonomists is younger in As and S.

Let us now consider the twelve questions that we posed above.

Which areas and countries publish more in zoological taxonomy?

The region that provided the highest number of publications is Europe, and then come, grouped together, North America, Asia and South America (Table 2-A). In Europe, no country provided more than 20% of publications, the most productive country being Germany (20%). In contrast, for the other main regions, a single country usually accounts for more than half the total publications: the USA for most publications from North America (76%), China for most publications from Asia (52%), Brazil for most publications from South America (62%), South Africa for most publications from Africa

Table 2 Relationships between studied regions and regions of researchers' labs

	A Studying regions (%)	B Studied regions (%)
Asia	20	25
South America	21	25
Multi-areas	-	12
North America	21	8
Oceania	7	8
Europe	29	7
Africa	2	6
Middle East	2	3
Central America	1	3
Antarctica	-	1
Without region	-	2

Studying regions: percentage of a fractional count of authors of publications. Studied regions: percentage of the number of publications. "Multi-areas" refers to work performed on more than one continent. Asia and South America are the most studied areas and belong to the areas with the highest numbers of publications.

(39%) and Turkey for most publications from Middle East (67%).

In which areas and countries are taxonomists working?

The number of publishing taxonomists per country is the highest for Europe (27%), then for South America (24%), North America (21%) and Asia (20%). Germany is the first country of Europe (20%); Brazil is the first country of South America (61%), the USA of North America (79%) and China of Asia (51%).

Which areas and countries mainly receive foreign taxonomists?

Concerning geographic mobility, 10% of the 427 authors who answered the question were not working in their country of nationality. Most were fully qualified researchers (69%), followed by predoctoral students (29%) and postdoctoral students (9%). Europeans are those who move the most (53%). Europe and North America are the areas with the highest rates of authors originating from other areas (37% and 25%, respectively). Unfortunately, our results on taxonomists' mobility do not allow us to know whether the weak mobility of taxonomists from Asia and South America to Europe, North America and Oceania is due to short-term mobility (which cannot be measured from our data) or to a more general lack of mobility.

How are international collaborations structured?

Collaborations are numerous: 73% of the publications have at least two authors. Three kinds of collaboration patterns were observed, corresponding to our previous

grouping in areas (Figure 5): regions which co-publish with every other area of the world (ENO); regions which mainly co-publish with themselves (AsS) or with ENO; and regions which do not publish much (AfM).

Authors from ENO area do not predominantly sign the publications as first authors (χ^2 test, $P = 0.054$).

The percentages of single-author articles is greater in ENO (28%) than in AsS (12%; χ^2 test, $P = 9.49e-08$).

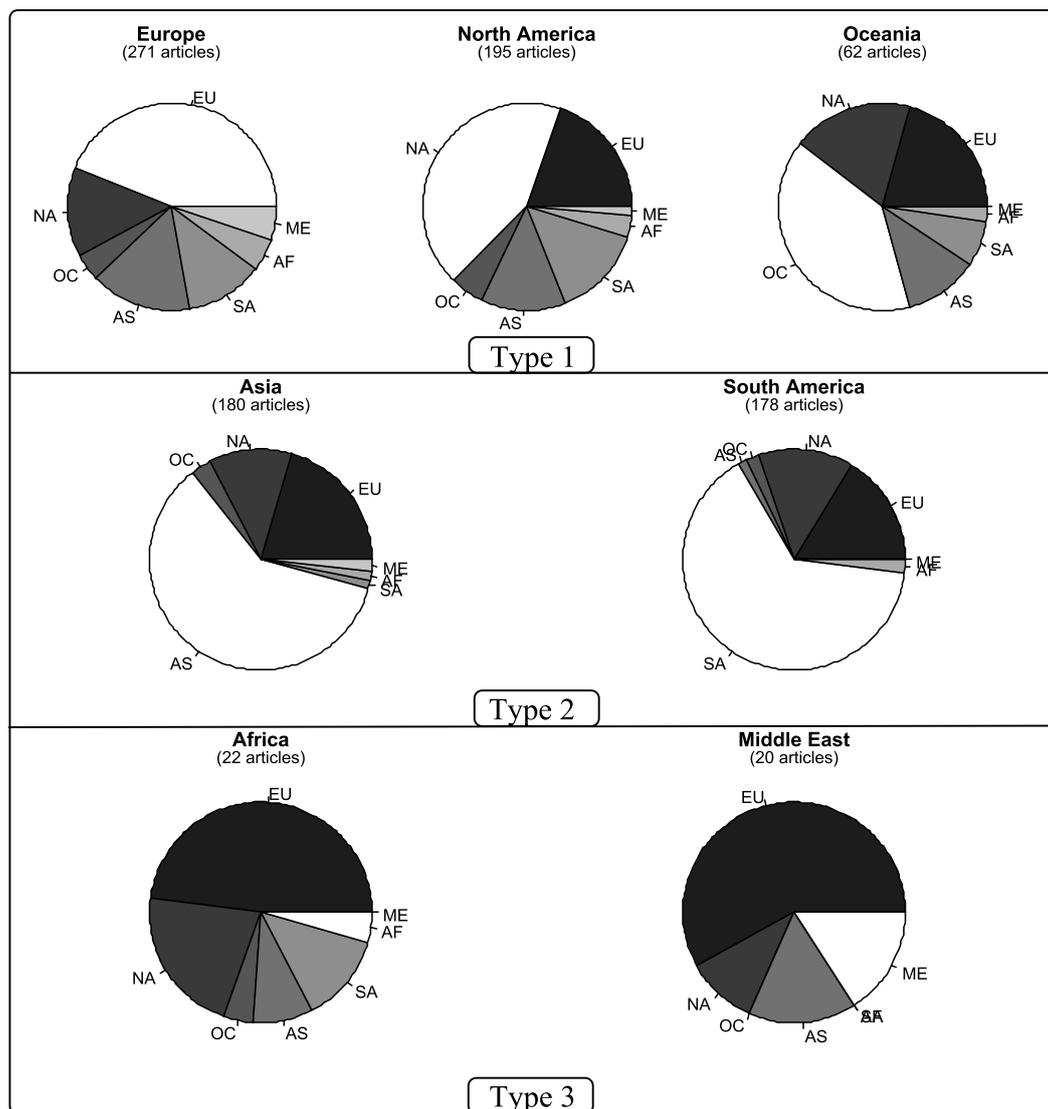
In which external areas and countries are new species described?

From a zootaxonomic point of view, Asia and South America are the most studied regions (see Table 2-B). Altogether they account for 50% of the publications and 61% of the new species. Latitudes are provided in the publications for 667 of the new species. Most of them (77%) come from latitudes higher than 10°, equally distributed between North and South. The remaining 23% are also equally distributed North and South (χ^2 : $P > 0.05$).

Figure 6 shows the result of the CA between the region of origin of the holotype and authors laboratories. The place of conservation of the holotype was used as supplementary data. Dimension 1 and dimension 2 together are strongly linked with both the origin of the holotypes and authors laboratories. They separate three groups: holotypes from AsS, holotypes from ENO and holotypes from AfM. AsS holotypes are linked to AsS collaborations and ENO-AsS collaborations whereas AfM holotypes are close to AfM, ENO-AfM, AsS-AM and ENO-AsS-AfM collaborations. However, AsS holotypes appear to be the category which is mainly studied by AS and ENO-AsS collaborations and AfM holotypes appear to be the category which is mainly studied by all areas of the world. The ENO holotypes category is linked to ENO collaborations and ENO holotypes appear to be studied by authors from ENO areas only. Thus, only ENO authors work on all areas of the world. Dimension 2 is linked with the atypical work on AM holotypes: authors from AfM are the only ones to work on their regions only, with numerous collaborations from all over the world.

Which areas and countries house the holotype specimens?

The supplementary data of Figure 6 show that holotypes are not always kept in the areas they come from. Authors from ENO area keep their holotypes as well as the holotypes from Africa and Middle East which they have studied in collaboration. On the other hand, authors from Asia and South America keep their holotypes even if studied in collaboration with authors from Europe, North America or Oceania.



Are museums preferred for carrying out taxonomic research and keeping the type specimens?

Author affiliations to universities are numerous (47%), followed by museums (28%), other institutions (21%) and private addresses (4%). Figure 7 shows the result of the CA between the geographical areas of authors and their type of institution. The places of conservation of the holotypes were used as supplementary data. Dimension 1 is strongly linked with geographical areas and separates the AsM area from the AfEO area. Universities and other institutions of work are linked to AsM and SN, whereas Private addresses and Museums are linked to AfEO. The supplementary data show that housing institutions differ according to the region and follow authors' affiliations,

except as concerns Universities. Holot_University is distant from auth_University on the map because a high number of museums are subordinated to universities (63% in our data), especially in AsM, and are thus not mentioned in the authors' address.

The proportion of museums in authors' affiliation decreases to 33% (AfEO), 3% (AsM) and 7% (SN) when the institution that subordinates the museum is taken into account. The proportion of universities also increases to 44% for AfEO, 70% for AsM and 67% for SN.

Noteworthy is the high proportion of museums (55%) and private collections (21%) for AfEO. 28% of the types kept in European private collections are holotypes and syntypes; 72% are paratypes.

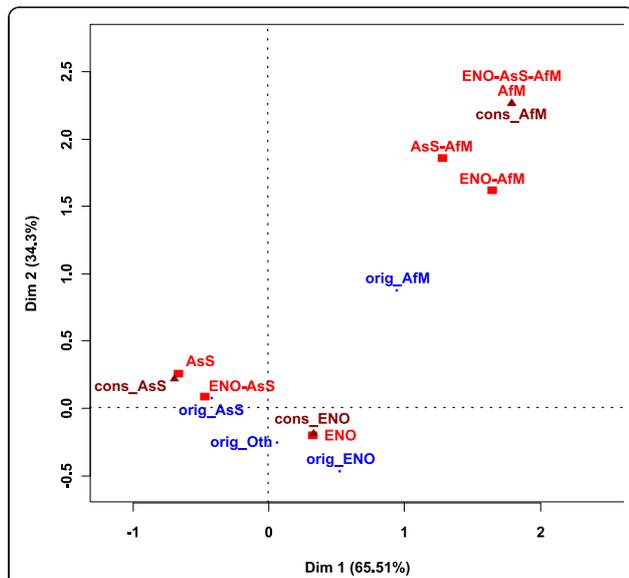


Figure 6 Result of the CA between the region of origin of the holotype and authors laboratory. The place of conservation of the holotype is used as supplementary data. *cons_*: region of conservation of the holotypes; *orig_*: origin of the holotypes. Af: Africa, As: Asia, E: Europe, M: Middle East, O: Oceania, N: North America, S: South America.

Are there more works on Vertebrata?

Most of the works on taxa (Table 3) deal with Arthropoda (69%), Chordata (16%), followed by Mollusca (3%) and Platyhelminthes (2%). Other taxa occur for less than 2% each and form the 10% remaining group. Arthropods

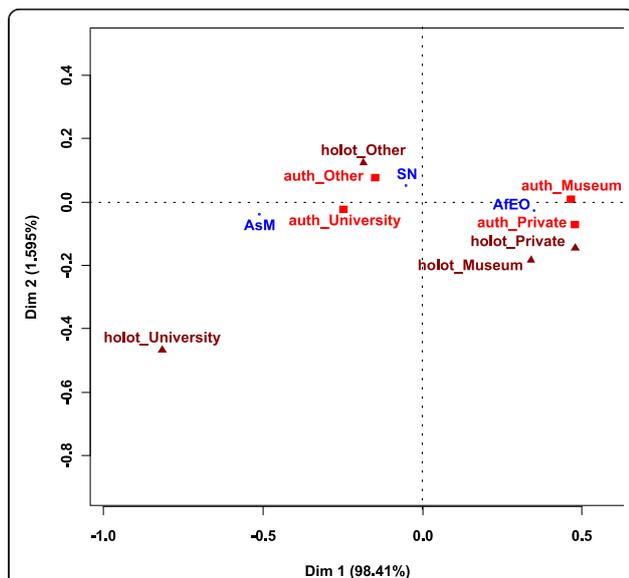


Figure 7 Result of the CA between the geographical areas of authors and their type of institution. The institution of conservation of the holotypes is used as supplementary data. *holot_*: holotypes' type of institution; *auth_*: authors' type of institution. Af: Africa, As: Asia, E: Europe, M: Middle East, O: Oceania, N: North America, S: South America.

Table 3 Proportion of publication and new species for each taxon

	% of publication	% of new species
Arthropoda	69	81
Insecta	45	58
Arachnida	10	11
Malacostraca	8	9
Other	6	3
Vertebrata	16	8
Amphibia	5	3
Actinopterygii	4	2
Reptilia	4	1
Chondrichthyes	1	1
Other	2	1
Mollusca	3	3
Platyhelminthes	2	2
Other	10	6
<i>Sum</i>	<i>100</i>	<i>100</i>

Left: n = 735. Right: n = 1416.

are mainly Insecta (64%) and Arachnida (14%), whereas Chordata are only represented by vertebrates.

It is thus not surprising that newly described species are mainly arthropods (81%) and vertebrates (8%). Many (51%) of the new species are described in taxonomic revisions, and only 35% in isolated works.

Figure 8 shows the result of the CA between studied phyla and the number of pages per new described species of each publication. The diversity of characters used to describe each new species and the number of authors of each document are used as supplementary data. Dimension 1 is strongly linked with phyla and separates the Vertebrata from the Arthropoda. Vertebrata is linked to category (10, 55] whereas Arthropoda is linked to categories with fewer pages per new species.

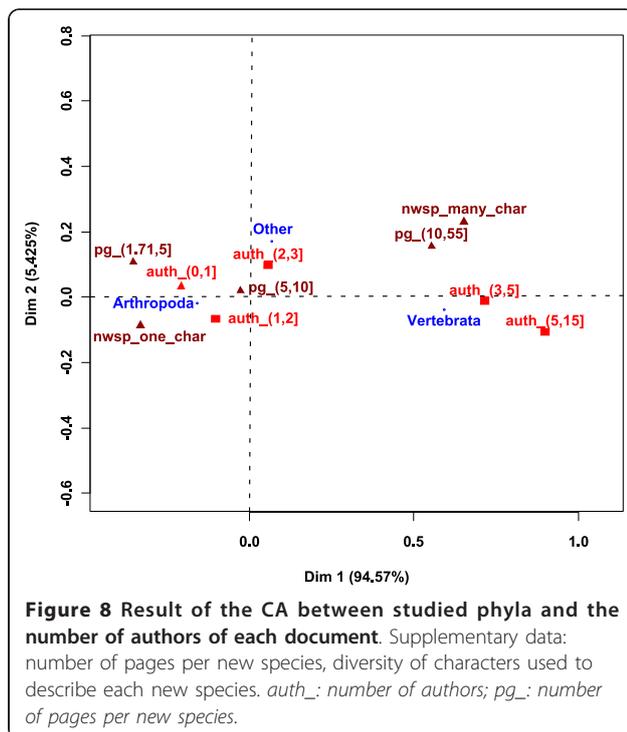
Are there more taxonomists involved in Vertebrata studies?

The supplementary data of Figure 8 show that the category Arthropoda is linked to few authors categories, whereas Vertebrata is linked to categories with more than three authors.

In the Vertebrata, a new species is described on average by 3.4 authors, whereas the figure is 0.9 in the Arthropoda. Thus, for an equal number of taxonomists, taxonomists working on arthropods describe more new species than taxonomists working on vertebrates.

Are there more characters involved in Vertebrata studies?

The supplementary data of Figure 8 show that the variety of characters used to describe the new species is higher for vertebrates. Whenever non-morphological



characters are used in Vertebrata descriptions, these are mainly molecular or bio-acoustic.

Are there more specimens involved in Vertebrata studies?

No significant difference between taxa was found in this respect. One species out of five was described from one specimen only (20%), or known from one locality only (21%). As much as 53% of the new species were collected from more than 2 localities. Two species out of five were described from specimen(s) of one sex only. Nine species out of 10 were described on the basis of a single stage of development (no significant difference between taxa). These figures vary when the year of collection is taken into account. New species known only from 1 specimen represent 64% of the new species with holotypes collected before the median year 1999. They represent 33% of the new species with holotypes collected after 1999 ($n = 1261$, $P = 5.83e-08$).

Discussion

Is zoological taxonomy still more developed in countries that have been historically taxonomically active?

This study of *Zootaxa*'s 2006-2007 publications supports the idea that zoological taxonomy is still more developed in countries that have historically been taxonomically active. Taxonomists from Europe, North America and Oceania are the only ones to collaborate largely with those from other regions, and they do so with all regions in the world. They also remain the only ones to

work on very wide geographical zones and to host researchers from abroad. A lack of studies in Africa has already been pointed out [11,25,43]. This lack still exists, as African regions are not well represented in our sample.

However, this study also shows that Asian and South American countries appear to play a prominent role in zoological taxonomy today. Their high number of publications, the fact that they keep their own specimens and the median age of their respondents show that they have a developed and active community of taxonomists. Previous studies on taxonomists' age could not be used for a rigorous comparison because the enquiries were not carried out at the same time [15,20,23,28,44-46].

Are these results artefacts created by our choice of the journal *Zootaxa*? We do not think so, for several reasons. First, we previously checked that our sampling was not biased questioning this respect. Second, the taxonomic growth of the two main countries responsible for these results (China and Brazil) has already been pointed out in the literature [10,47]. China's funding for research increased by 30% each year between 1998 and 2006, which benefited taxonomy: "The evaluation of the funded projects in the Division [of Zoology] shows that animal morphology and taxonomy are the main supported areas, accounting for half of the total funded" [48]. In 2004, China became the fifth country in terms of number of publications. The number of students has been increasing everywhere in this country [49]. An analysis of Brazil's scientific production carried out by The Observatoire des Sciences et Techniques (Paris, France) showed that it was quickly growing [50,51]. The figures were even underestimated: a bibliometric study [50] showed that many Brazilian articles are published in Portuguese, in national journals. Thus they were not included in the calculation. Moreover, a recent study on amphibian taxonomic effort and expertise supports these results concerning China and Brazil [52]. It also found that the continental distribution of authors in *Zootaxa* ($n = 5663$) was similar to the continental distribution of authors working on a given taxon (i.e. Amphibians, $n = 647$).

Another argument in favour of this idea involves the geographical origin of the new species. Surprisingly, no latitude preference was identified regarding the origin of the newly described species. This contradicts the widespread idea that latitudes close to the equator have a richer biodiversity [53]. Why, then, do our data fail to find high proportion of new species in the belt of 10° above and below the equator?

In this respect we found that places where the highest numbers of species were described recently (China, Brazil) are also those with the highest number of taxonomists. A correlation exists between the number of

taxonomists and the number of species known in a phylum [54]. It is thus likely that the high number of newly described species in China and Brazil is a reflection of the *taxonomic involvement* of these countries, in terms of manpower and funding. Thus, if China and Brazil appear to be the most studied regions, this is not only because they are large countries with a rich biodiversity [55-57], but also because they have numerous taxonomists and have decided to invest in this branch of research.

Finally, it is striking that a high proportion of our respondents were professionals (fully qualified researchers, students) despite the high number of amateurs in entomology [9]. *Zootaxa* would therefore reflect a biased sub-sample of the taxonomic community, including mostly professionals. One may wonder whether part of this phenomenon might be due to the fact that some amateur taxonomists, especially in entomology, tend to publish their works in their national languages in national periodicals whenever they still exist [58]. Unfortunately, professional categories are not expressed the same way between studies, hence our figures are hardly comparable with previous works [15,24,28,54]. We also have to keep in mind that the authors who provided their email addresses are mainly the corresponding authors - a factor that might influence our results.

Is zoological taxonomy still in favour of vertebrates, vs. arthropods?

On the basis of taxonomic output (number of publications, number of new species), arthropods, especially insects, appear to be much more studied than vertebrates. Vertebrates nevertheless concentrate a higher number of taxonomists, characters and methods used for each new species. Thus, even if more results are produced on arthropods, their study involves a lesser amount of means (number of taxonomists, number and diversity of methods used).

The long-studied vertebrates [2,59] are nowadays studied with modern methods. Most of the time, these methods complement the morphological ones. As a consequence, morphological studies are a necessary step in any taxonomic work. They are not a sign of intellectual or technological backwardness when presented alone [60]. They are a strong indication of the amplitude of our ignorance of the taxon in question. The data we gathered on the numbers of specimens, localities and kinds of characters used to recognize and describe a new taxon clearly show that we are very far from "knowing" a species once it has been "described" and named [61].

Moreover, the taxa which benefit from a high taxonomic involvement have nevertheless the lowest rates of description of new species. If we aim at increasing the number of known species as quickly as possible,

taxonomists should give priority to taxa which do not require these new methods.

We do not think that these results are artefacts created by our choice to work on the journal *Zootaxa*. We previously checked that our sampling was not biased for this question and we already know that vertebrates have been studied for a longer time [2,59]. Our observations match this idea.

This disequilibrium of the taxonomic effort in favour of vertebrates is problematic if we aim at better knowing that which we know the least. Insects represent 58% of the new described species and vertebrates 7.2%. Yet, according to some estimates (figures computed from the sum of the 'World Descr.' and 'Estimate World' figures for chordates and invertebrates in [56]), vertebrates would only represent 0.3% of the numbers of species that remain to be described, whereas insects represent 74%. Thus, this distribution of taxonomic effort between taxa is unsuitable when considering the needs [16,23,24,54,62]. The international and national scientific communities should hire more people to focus on arthropods, especially entomologists.

The nature of the works published in *Zootaxa* show that this journal gives priority to alpha-taxonomy. "*Zootaxa considers papers on all animal taxa, both living and fossil, and especially encourages descriptions of new taxa*" [63]. We therefore consider our main results valid for zoological alpha-taxonomy.

Conclusions

We found that the disequilibrium does not correspond exactly to that usually admitted. The USA, China and Brazil are currently the three leading countries in zoological alpha-taxonomy. However, each of them presents a different pattern. Taxonomists from Asia and South America are younger and mainly work in universities, rather than museums. A bias in favour of vertebrates still exists if we refer to the effort invested in each group to produce taxonomic data, but not for the number of papers. Finally, we insist on the idea that "describing a species" is very different from "knowing a species". The taxonomic involvement of a country, in terms of manpower and funding, appears here to be a key factor in the development of fruitful taxonomic research. If we aim at increasing our knowledge of the species of the world, we still have much efforts to make in terms of the efficient distribution of the taxonomic effort between taxa.

Acknowledgements

We thank all those who took time out of their already busy schedules to fill in and return the questionnaire. Thanks to Zhang Zhi-Qiang, who encouraged this study, provided free access to *Zootaxa's* articles and provided critical comments on the work. We would also like to thank

Sébastien Ballesteros, Julien Barnier and Gaël Millot for their statistical help. If any statistical mistake can be found in this paper, it must nevertheless be attributed to the authors. We would like to thank Sandra Dos Santos for her translations of the Portuguese papers and Annie Amirda, Julie Delacôte and Mark Judson for the English corrections. We would also like to thank Philippe Grandcolas, Aurélien Miralles, Antoine T. Mantilleri, Ferdinando Boero, Mark Judson and four anonymous referees for their critical comments.

Author details

¹UMR 7205 OSEB, Département Systématique et Evolution, Muséum national d'Histoire naturelle, CP 30, 25 rue Cuvier, 75231 Paris Cedex 05, France.
²SAP2S-DASTR, CNRS, 3 rue Michel Ange, 75794 Paris Cedex 16, France.

Authors' contributions

ET designed the sampling procedure and the database, compiled the database, participated in the statistical analysis and drafted the manuscript. CB designed and participated in the statistical analysis. AS participated in the analyses of the results and helped to draft the manuscript. AD designed the study, participated in the analysis of the results, and helped to draft and translate the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 19 October 2010 Accepted: 18 March 2011

Published: 18 March 2011

References

- Wilson JRJ, Proches S, Braschler B, Dixon ES, Richardson DM: **The (bio) diversity of science reflects the interests of society.** *Frontiers in ecology and the environment* 2007, **5**:409-414.
- Proença VM, Pereira HM, Vicente L: **Organismal complexity is an indicator of species existence value.** *Frontiers in ecology and the environment* 2008, **6**:298-299.
- Cribb TH: **Living on the edge: parasite taxonomy in Australia.** *International Journal for Parasitology* 2004, **34**:117-123.
- Bonnet X, Shine R, Lourdaïs O: **Taxonomic chauvinism.** *Trends in Ecology & Evolution* 2002, **17**:1-3.
- GBIF and Data Repatriation to Countries of Origin. [<http://www.gbif.org/index.php?id=488>].
- EDIT: **Report and online catalogue on existing taxonomic training resources in Europe.** Brussels: Royal Belgian Institute of Natural Sciences; 2006.
- Results of Surveys of Taxonomic Expertise in Parasitology in Canada and of Interest in National Faunal Projects on Parasites in Canada.** [<http://www.biology.ualberta.ca/parasites/ParSec/texten/modgap.htm>].
- Pallini A, Fadini MAM, Venzon M, De Moraes GJ, Barros-Battesti DM: **Demandas e perspectivas para a Acarologia no Brasil (Demands and perspectives to acarology in Brazil).** *Neotropical Biology and Conservation* 2007, **2**:169-175.
- Löbl I, Leschen RAB: **Demography of Coleopterists and Their Thoughts on DNA Barcoding and the Phylocode, with Commentary.** *The Coleopterists Bulletin* 2005, **59**:284-292.
- Eggleton P: **Termite species description rates and the state of termite taxonomy.** *Insectes Sociaux* 1999, **46**:1-5.
- Green SV: **The taxonomic impediment in orthopteran research and conservation.** *Journal of Insect Conservation* 1998, **2**:151-159.
- House of Lords Science and Technology Committee: *Systematics and Taxonomy: Follow-up. 5th report.* London 2008.
- Valdecasas AG, Bello E, Becerra JM: **La Taxonomía en España.** Arbor: Ciencia, Pensamiento y Cultura; 1992:**561**:77-102.
- Tillier S, De Wever P, (Eds): **Systématique: ordonner la diversité du vivant - Rapport sur la science et la technologie n°11.** Paris: Editions Tec & Doc; 2000.
- Australian Government department of the Environment Water Heritage and the Arts: *Survey of Australian Taxonomic Capacity* 2006.
- Marques AC, Lamas CJ: **Sistemática zoológica no Brasil: estado da arte, expectativas e sugestões de ações futuras.** *Papéis Avulsos de Zoologia* 2006, **46**:139-174.
- Kraus O, (Ed): **Zoologischer Systematik in Mitteleuropa (Zoological systematics in Central Europe).** Hamburg & Berlin: Paul Parey; 1976.
- Schminke HK: **Wiederaufbau von Forschung und Lehre in Zoosystematik - eine nationale Aufgabe.** *Spektrum der Wissenschaft* 1994, **9**:114-116.
- Smith GF, Buys M, Walters M, Herbert D, Hamer M: **Taxonomic research in South Africa: the state of the discipline.** *South African Journal of Science* 2008, **140**:254-256.
- Simonetti JA: **Biodiversity and a taxonomy of Chilean taxonomists.** *Biodiversity and Conservation* 1997, **6**:633-637.
- Salazar Vallejo SL, Gonzalez NE, Schwindt E: **Taxonomia de invertebrados marinos: necesidades en latinoamérica (Marine invertebrate taxonomy: needs in Latin America).** *Interciencia* 2008, **33**:510-517.
- Blackmore S, Cutler DF, (Eds): **Systematics Agenda 2000: the challenge for Europe.** Cardigan: Samara Publishing for the Linnean Society; 1996.
- Shimura J, (Ed.): **Global Taxonomy Initiative in Asia. Research Report from the National Institute for Environmental Studies.** 2003.
- Gaston KJ, May RM: **Taxonomy of taxonomists.** *Nature* 1992, **356**:281-282.
- Klopper RR, Smith GF, Chikuni AC: **The Global Taxonomy Initiative in Africa.** *Taxon* 2002, **51**:159-165.
- Zhang ZQ: **The first five years of Zootaxa.** *Zootaxa* 2006, **1111**:68.
- Top journals for new taxa and top journals for new taxonomic/nomenclatural acts.** [<http://www.organismnames.com/metrics.htm?page=tsj>].
- Betsch J-M, (Ed): **Livre Blanc de la Systématique.** Société Française de Systématique; 1989.
- Geeta R, Levy A, Hoch JM, Mark M: **Where do most taxonomists work, and does it matter? Poster.** *Annual AIBS conference on "Bioethics in a changing world"; Washington, D. C., USA* 2003.
- Tillier S: **Une espèce menacée: les taxonomistes en France.** In *Systématique: ordonner la diversité du vivant - Rapport sur la science et la technologie n°11.* Edited by: Tillier S, de Wever P. Paris: Editions Tec 2000:20-23.
- Zootaxa website.** [<http://www.mapress.com/zootaxa/>].
- R Development Core Team: *R: a language and environment for statistical computing.* Vienna, Austria 2008.
- CompressWeb: **Email Questionnaire 4.1.** 2007.
- Central Intelligence Agency: **The World Factbook 2008.** Washington, DC: Central Intelligence Agency; 2008.
- Dray S, Dufour A, Chessel D: **The ade4 package-II: Two-table and K-table methods.** *R News* 2007, **7**:47-52.
- Ripley BD, Lapsley M: **RODBC: ODBC Database Access. R package version 1.2-3.** 2008.
- Meyer D, Zeileis A, Hornik K: **The Strucplot Framework: Visualizing Multi-Way Contingency Tables with vcd.** *Journal of Statistical Software* 2006, **17**:1-48.
- Zeileis A, Meyer D, Hornik K: **Residual-based Shadings for Visualizing (Conditional) Independence.** *Journal of Computational and Graphical Statistics* 2007, **16**:507-525.
- Husson F, Lê S, Pagès J: *Analyse de données avec R* Rennes: Presses Universitaires de Rennes; 2009.
- Benzecri JP: In *L'analyse des correspondances. L'analyse des données. Volume 2.* Paris: Dunod; 1973.
- Mayr E, Ashlock PD: *Principles of systematic zoology.* New-York, Second 1991.
- Dubois A: **Drôles d'espèces. Hybridation, perturbations de la méiose et spéciation dans le règne animal: quelques points délicats de terminologie, d'éidonomie et de nomenclature.** In *Peut-on classer le vivant? Linné et la systématique aujourd'hui.* Edited by: Prat D, Raynal-Roques A, Roguenant A. Paris: Belin; 2008:169-202.
- Slater JA: **The place of systematists in contemporary society.** *Bulletin of the Entomological Society of America* 1984, **30**:4-7.
- House of Lords Science and Technology Committee: *Systematic Biology Research. 1st report 1991-92.* London 1992.
- Krebs JR, (Ed): **Evolution and Biodiversity: The New Taxonomy.** London: UK Natural Environment Research Council; 1992.
- Tillier S: **Les communautés scientifiques nationales.** In *Systématique: ordonner la diversité du vivant - Rapport sur la science et la technologie n°11.* Edited by: Tillier S, de Wever P. Paris: Editions Tec 2000:42-43.
- Joly CA: **Taxonomy: programmes developing in the South too.** *Nature* 2006, **440**:24-24.
- Division of Zoology, Animal Husbandry and Veterinary Science and Aquatic Science Animal Husbandry and Veterinary Science.** [http://www.nsf.gov.cn/e_nsf/desktop/zn/jn/0124.htm].

49. Observatoire des Sciences et des Techniques: *Chine: dossier-pays OST, décembre 2006* 2006.
50. Observatoire des Sciences et des Techniques: *Brésil: dossier-pays OST, juin 2003* 2003.
51. Michán L, Llorente Bousquets J: **Bibliometría de la sistemática biológica sobre América Latina durante el siglo XX en tres bases de datos mundiales.** *Revista de Biología Tropical* 2009, **58**:531-545.
52. Rodrigues ASL, Gray CL, Crowter BJ, Ewers RM, Stuart SN, Whitten T, Manica A: **A Global Assessment of Amphibian Taxonomic Effort and Expertise.** *Bioscience* 2010, **60**:798-806.
53. Dubois A: **Threatened amphibians.** *Alytes* 2009, **27**:25-37.
54. Bello E, Becerra JMG, Valdecasas A: **Counting on taxonomy.** *Nature* 1992, **357**:531.
55. Biodiversity Committee of the Chinese Academy of Sciences: **Species diversity.** *Biodiversity in China: status and conservation needs.* Beijing 1992.
56. Chapman AD: **Numbers of Living Species in Australia and the World.** Canberra: Australian Biodiversity Information Services, 2 2009.
57. Lewinsohn TM, Prado PI: **How Many Species Are There in Brazil.** *Conservation Biology* 2005, **19**:619-624.
58. Dubois A: **Should we care about taxonomic journals? A plea for Biotaxa, a new kind of online subscription package.** *Newsletter (European Distributed Institute of Taxonomy)* 2008, **9**:12-14.
59. California Academy of Sciences: *A century of progress in the natural sciences, 1853-1953.* San Francisco 1955.
60. La Salle J, Wheeler Q, Jackway P, Winterton S, Hobern D, Lovell D: **Accelerating taxonomic discovery through automated character extraction.** *Zootaxa* 2009, **2217**:43-55.
61. Dubois A: **List of European species of amphibians and reptiles: will we soon be reaching stability?** *Amphibia-Reptilia* 1998, **19**:1-28.
62. Blackmore S: **A Summary of UK Based Expertise in Systematic Biology.** UK SYSTEMATICS FORUM; 1996.
63. **Information for authors.** [<http://www.mapress.com/zootaxa/support/author.html>].

doi:10.1186/1742-9994-8-5

Cite this article as: Tancoigne *et al.*: **Insights from Zootaxa on potential trends in zoological taxonomic activity.** *Frontiers in Zoology* 2011 **8**:5.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Annexe B

Détail des champs du *ZR*

NOM	DESCRIPTION
FN = File Name	-
VR = Version Number	-
PT = Publication Type	-
AN = Zoological Record Accession Number	A unique identification number for each record in the database
DT = Document Type	The selected document type. In database years 1978-1997, the document types are Article, Book, and Book Chapter. In database year 1997, three more document types were introduced : Meeting, Meeting paper, and Software
TI = Title	The English-language title of the source article or book chapter. If the title appears within square brackets, it has been translated from another language
FT = Foreign Title	The title of the source publication in the original language (if the original language is not English)

... à suivre ...

TABLE B.1 – suite

NOM	DESCRIPTION
AU = Authors	<p>The author(s) of the source document. Author names appear as last name followed by a first name and/or initials. This field also contains the names of corporate authors.</p> <p>All author names from the source document are provided in records from 1994 to the present. Prior to 1994, up to ten authors are shown, with "et al" in the tenth position.</p> <p>Records of books with no author(s) will display the word "Anonymous" in the Author field. A record of a collection of meeting abstracts will display "Abstracts" in the Author field.</p>
BA = Book Authors	<p>The author(s) of the book or book chapter.</p> <p>Records of books with no author(s) will display the word "Anonymous" in the Author field. A record of a collection of meeting abstracts will display "Abstracts" in the Author field</p>
ED = Editors	<p>The name of one or more editors of a source publication. Be aware that a corporate author can also be an editor</p>
SO = Source	<p>The title of the book or journal. Source information may include : Volume, Issue, Part, Supplement, Page range, Date, Publication Date (cover date), DOI (Digital Object Identifier)</p>
VL = Volume	-
IS = Issue	-
SU = Supplement	-
PS = Page Span	-
PD = Publication Date	-
PY = Year Published	-
UR = Journal URL	-
AW = Item URL	-
NT = Notes	<p>The Notes field contains miscellaneous bibliographic information such as translation information</p>

... à suivre ...

TABLE B.1 – suite

NOM	DESCRIPTION
LA = Language	The selected language of the source document. Note : In some product databases, such as Food Science and Technology Abstracts, the field label may also say Language of Summary
SL = Summary Language	The language(s) of the summary or abstract. The source document may include multiple abstracts translated from the original language
ME = Medium	The medium in which the source document was published
AB = Abstract	A brief summary or description of the essential content from the source document
C1 = Author Address	The address of the author who should receive communications about the source publication such as requests for reprints
EM = E-mail Address	The e-mail address of the main author or the author who should receive communications about the source publication
PU = Publisher	The publisher of the source document. Includes the publisher name and publisher address
PA = Publisher Address	-
SN = ISSN	The International Standard Serial Number is a unique eight-digit number that identifies periodical publications such as journals and electronic publications
BN = ISBN	The International Standard Book Number is a unique machine-readable identifier that marks a book unmistakably
BD = Broad Terms	Controlled terms from the following hierarchies of the Zoological Record Thesaurus : Subject, Geographical, Palaeontological, Systematic, Taxa Notes. These are broader terms of the controlled terms used to index the source document

... à suivre ...

TABLE B.1 – suite

NOM	DESCRIPTION
DE = Descriptors Data	<p>Controlled and non-controlled terms used to index the source document. Descriptors Data are displayed in a table. Each row of the table contains one or more animal names and related subject index terms.</p> <p>The Organism column contains the name(s) of animals discussed in the source document.</p> <p>The Controlled Term column contains terms from the Subject, Geographical and Palaeontological hierarchies in the Zoological Record Thesaurus.</p> <p>The Modifier column contains free-text or non-controlled indexing terms</p>
TN = Taxa Notes	<p>Controlled terms from the Taxa Notes hierarchy of the Zoological Record Thesaurus. Taxa Notes terms are English vernacular names for broad animal groups for animals discussed in the source document</p>
ST = Super Taxa	<p>Controlled terms from the Super Taxa hierarchy of the Zoological Record Thesaurus. Super Taxa terms are scientific names displayed to show the taxonomic hierarchy to which animals discussed in the source document belong</p>
OR = Systematics	<p>Controlled and non-controlled terms relating to animal names, systematics, and taxonomy which are used to index the source document. Systematics Data are displayed in a table. Each row of the table contains one animal name and related index terms</p> <p>The Classifier column contains the appropriate controlled term from the Super Taxa hierarchy of the Zoological Record Thesaurus that relates to the Organism Name</p> <p>The Organism Name column contains the animal name</p> <p>The Author/Year column contains the name of the author(s) and year of publication of the animal name.</p> <p>The Controlled Term column contains terms from the Systematics hierarchy of the Zoological Record Subject Thesaurus</p> <p>The Modifier column contains free-text or non-controlled indexing terms</p>

... à suivre ...

ANNEXE B : Détail des champs du *ZR*

TABLE B.1 – suite

NOM	DESCRIPTION
UT = Thomson Reuters Unique Article Identifier	-
ER = [End of Record]	-
EF = [End of File]	-

Annexe C

Exemples de scripts de traitement des données sous R

C.1 Fonction d'import des notices

```
#les noms de fichier à importer sont de la forme "name_1-500.txt"
#la fonction prend en entrée
  #le chemin du répertoire contenant les fichiers
    #la partie commune aux noms de fichiers
    #le nombre total de références du corpus
#elle renvoie une liste contenant
  #le dataframe d'importation
  #les références des lignes contenant des erreurs d'importation

function(directory , name, number){

files_number <- trunc(number/500,0) # nombre de fichiers - 1

#import des données dans 'import'
import <- read.delim(paste(directory , name, "_1-500.txt", sep=""), quote="",
                    header=T, sep = "\t")

for (i in seq(1,files_number-1,1)){
  #attention car si on met 1:files_number-1, i commence à 0
  temp <- read.delim(paste(directory , name, "_", 500*i+1, "-",
                        500*i+500, ".txt", sep=""),
                    quote="", header=T, sep = "\t")
  import <- rbind(import , temp)
```

```

}
import <- rbind(import, read.delim(paste(directory, name, "_",
                                     500*files_number+1, "-", number, ".txt", sep=""),
                                     quote="", header=T, sep = "\t"))

#transformation de 'import' en matrice pour rendre ses valeurs modifiables
import <- as.matrix(import)
#suppression des valeurs manquantes
import[which(is.na(import))] <- ""

#recherche des lignes décalées (i.e. ne possédant pas "ZOOR" dans AN)
errors <- grep("ZOOR", import[, "AN"], invert = T)

return(list(import=import, errors=errors))
}

```

C.2 Fonction de déduplication des notices

```

#la fonction prend en argument
#un dataframe de références importées
#dont on veut enlever les duplicatas
#elle renvoie
#le même dataframe
#sans les duplicatas (on conserve la référence la plus récente)

function(data){
#appel des packages nécessaires
tcltk <- require(tcltk)
if (!tcltk)
  stop("The 'tcltk' package is absent")

#repérage des références avec les mêmes
#auteur(s), titre, source, année de publication
duplicates <- duplicated(data[, c("AU", "TI", "SO", "PY")])
#duplicates[1:10]
#[1] FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
if("TRUE" %in% duplicates){
#on ne conserve pas les duplicats intitulés "[Title unknown.]"
titles <- which(data[, "TI"] != "[Title unknown.]")
duplicates <- intersect(titles, which(duplicates==T))
}
}

```

ANNEXE C : Exemples de scripts de traitement des données sous **R**

```
#pour chaque paire, on conserve la référence la plus récente
#elle est repérée d'après le numéro d'accension ZR
toRemove <- vector()#vecteur avec les références des lignes à supprimer
for (i in 1:length(duplicates)){#pour chaque paire
  #on extrait le numéro d'accension ZR du champ "AN"
  pair <- data[which(data[, "TI"] == data[duplicates[i], "TI"]
    & data[, "PY"] == data[duplicates[i], "PY"]
    & data[, "AU"] == data[duplicates[i], "AU"]
    & data[, "SO"] == data[duplicates[i], "SO"]), "AN"]
  pair <- sort(pair, decreasing=F)
  #on ajoute à toRemove le(s) numéro de(s) ligne(s) la plus ancienne(s)
  toRemove <- c(toRemove, as.numeric(names(pair[-length(pair)])))
}
toRemove <- unique(toRemove)
data <- data[-toRemove,]
#affichage du nombre de références supprimées
ReturnVal <- tkmessageBox(message = paste(length(toRemove),
  "_articles_supprimés"), icon="info", type="ok")
}

else{
ReturnVal <- tkmessageBox(message = "Pas_de_doublons", icon="info",
  type="ok")
}

return(data)
}
```


Annexe D

Définitions de la taxonomie dans la littérature

D.1 Quelques définitions des termes *taxinomie*, *taxonomie* ou *taxonomy*

- (1) [the word taxonomy] *denotes the science of arrangement or system, and hence classification* (Hall et Clements, 1923)
- (2) *Taxonomy attempts to isolate from the actual experience a sort of a photographic image of the living world, and to give a comprehensible description of the regularities observed in this image* (Dobzhansky, 1936, p. 353)
- (3) *Taxonomy deals with the descriptions and illustrations of all the prime groups [the genus and the species, see p. 515] of the animal kingdom* (Metcalf, 1937, p. 514)
- (4) *systems of classification* (Thorpe, 1940)
- (5) *to place before the other workers in biology as full and complete a record of the forms living in the world as is possible with our present support and opportunities* (Metcalf, 1946)
- (6) *taxonomy, the formal nomenclature and description of organisms* (Anonymous, 1946a)
- (7) [understanding] *the immense floras and faunas of the world* (Anonymous, 1946b)
- (8) [...] *for the work on species and their segregates the term taxonomy is available. It is with taxonomy alone in this sense that we are concerned here-the segregation, naming, and identification of genera and species and lesser groups* (Blackwelder, 1959)

- (9) *Taxonomy is the theoretical study of classification, including its bases, principles, procedures, and rules* (Simpson, 1961, p. 11)
- (10) *One aim of taxonomy is to produce a system of order which will reflect as nearly as possible the evolutionary relationships between the entities being classified* (Stebbins, 1969, p. 26)
- (11) *determining the composition of the community in terms of organisms and their life histories and dispersal powers, and so forth* (Evans, 1972)
- (12) *Modern taxonomy is no longer the science of "giving names and nomenclature"; the final result of taxonomy [...] has to be the understanding of the species concept, the genus concept, the family concept, and so forth, and the understanding of the relation and interaction between species or between genera* (Van der Spoel *et al.*, 1972)
- (13) *record information about diversity on a continental or global scale* (Wilson, 1972)
- (14) *the orderly classification of organisms according to their presumed natural relationships* (Richardson, 1984)
- (15) *Systematics has two branches, microtaxonomy [...] taxonomy at the level of the species, and [...] macrotaxonomy, the classification of birds* (Mayr, 1989)
- (16) [it deals] *with variation, relationships and phylogeny* (Bremer *et al.*, 1990)
- (17) *classification and description of [species]* (Harley, 1990)
- (18) *Taxonomy is the theory and practice of classifying organisms* (Mayr et Ashlock, 1991)
- (19) *taxonomy proper, the discipline concerned with methods for communicating the results of systematic analysis* (de Queiroz et Gauthier, 1992, p. 450)
- (20) *Here we define taxonomy as the science of classifying organisms. Classification involves more than just cataloguing and naming : it involves elucidating the evolutionary relationships among organisms, as well as problems related to the diversity of organisms alive today and at different times in the past* (Krebs, 1992)
- (21) *Taxonomy in the strict sense refers to all information science aspects of handling the different sets of organisms* (Anonymous, 1995)
- (22) [...] *on peut ainsi réduire la définition de la taxonomie au sens strict à la science de la description, de la dénomination et de l'identification des taxons, alors que la biosystématique est pratiquement confondue avec la classification phylogénétique et ses explications évolutives* (Tillier et De Wever, 2000)

ANNEXE D : Définitions de la taxonomie dans la littérature

- (23) [...] *taxonomy as the practice of describing and classifying organisms* (Coomans, 2000)
- (24) *Taxonomy, the classification of living things* (Godfray, 2002a)
- (25) *Taxonomy provides a vocabulary to discuss the world : specifically, the names of organisms* (Knapp et al., 2002)
- (26) *taxonomy and systematics, the naming and classification of living things,* (Thiele et Yeates, 2002)
- (27) *science of discovering, describing and naming individual species of fauna and flora of the world, and aims to explain their relationships in order to provide a predictive classification system* (Klopper et al., 2002)
- (28) *The purpose of taxonomy is the identification of species and their assignment to higher level taxa* (Tautz et al., 2003, p. 72)
- (29) *Taxonomy, the scientific study and practice of classification* (Wilson (2003, p. 78), Crisci (2006b, p. 217))
- (30) *Taxonomy is the discipline of biology whose aim is to allocate all living organisms to formal classificatory units (the taxa) referred to a hierarchy of categories (e.g., species, genera, families, etc.) and to name all these units according to an international nomenclatural system of Latin nomina, which constitutes a language common to all biologists* (Dubois, 2003b)
- (31) *Taxonomy, a term that we usually restrict to the classification and naming of organisms [...]* (Raven, 2004)
- (32) *taxonomy can justly be called the pioneering exploration of life on a little known planet* (Wilson, 2004)
- (33) *packaging species-level information for users, and enable us all to read wild biodiversity* (Janzen, 2004)
- (34) *Taxonomy [...] establishes the characteristics, relationships, range and identity of the components of biological systems* (Mather, 2004)
- (35) *describe [...] diversity* (Sheldon et al., 2005)
- (36) *Defining the biodiversity of our planet* (McCallum et McCallum, 2006)
- (37) *expertise in the identification of [...] species* (Costello et al., 2006)
- (38) *discover and describe Earth's species ; classify species according to their phylogenetic relationships ; and make what is known about them accessible through Linnaean classifications and names* (Wheeler, 2007; de Carvalho et al., 2008)
- (39) *[...] the three principal tasks of taxonomy are description, identification and phylogeny* (Knapp, 2007)

- (40) *Taxonomy is the scientific discipline of describing, delimiting and naming organisms* (House of Lords Science and Technology Committee, 2008)
- (41) *The study will cover all three major elements of taxonomy identified in the report, identification, classification and phylogeny* (Anonymous, 2008a)
- (42) *discovery and documentation of our biological diversity* (Anonymous, 2008b)
- (43) *taxonomy—the naming, description and identification of organisms* (Bateman, 2008)
- (44) *Taxonomy is the science of the description and classification of organisms* (Guerra-Garcia *et al.*, 2008)
- (45) *The object of interest of taxonomy is the almost unimaginably diversified world of living creatures : the studies comprise the entire process from meticulous analysis of intra- and interpopulational relations to the, resisting any schematic approach, final synthesis in form of natural classification* (Hołyński, 2008)
- (46) *La taxonomía o sistemática (aquí se emplean ambos términos indistintamente) es la subdisciplina de la biología encargada de la descripción, nomenclatura, clasificación, teoría e historia de la clasificación de los seres vivos* (Wiley, 1981) (Michán *et al.*, 2008)
- (47) *biodiversity research* (Prathapan *et al.*, 2008)
- (48) *permite la delimitación de las especies de manera estable, consistente y explícita* (Mace, 2004) (Salazar Vallejo *et al.*, 2008)
- (49) *Taxonomy is the science of detecting, describing, naming and classifying organisms* (Smith *et al.*, 2008)
- (50) *identifying and naming biodiversity* (Szwedo *et Iwan*, 2008)
- (51) *Taxonomy is the geometry of nature* (Beeton, 2008)
- (52) *contemporary taxonomy [...] documents and publicizes the quickly disappearing wonder of biological diversity* (Hulcr, 2009)
- (53) *describe and name the surrounding biodiversity* (La Salle *et al.*, 2009)
- (54) *The discipline aimed at naming species, i.e., taxonomy* (Boero, 2010)
- (55) *Taxonomy comprises the theory and practice of describing, naming, and ordering groups of organisms termed taxa* (Wiley *et Lieberman*, 2011, p. 8)

D.2 Quelques définitions des termes *systematique* ou *systematics*

- (56) *systematic science - in the discrimination of species and varieties, in the study of the relations of these groups to each other and to their environment (Jordan, 1888)*
- (57) *systematic botany as represented by the classification of species (Hitchcock, 1916)*
- (58) *Theoretically the work of the systematist is to build a classification of living organisms that will show –assuming evolution to be a fact–their natural relationships and enable others to be fitted into the scheme when they are subsequently found. In practice his work may be divided into, firstly, the general work of identification [...] and, secondly, the prosecution of fundamental research in phylogeny and zoogeography (Smart, 1940)*
- (59) *systematics plays a double role, viz., the defining, describing and naming of groups, with the inference that in arranging them their descent, and hence their genetic relationship, is being portrayed at the same time (Cole, 1941)*
- (60) *Systematics, the study of biological diversity (Wilson, 1985b)*
- (61) *Systematics is not simply the identification of specimens, it is the reconstruction of evolutionary history (O’Hara et al., 1988)*
- (62) *systematic biology, the classification of living things based on evolutionary relationships (Gee, 1992)*
- (63) *Nommer et classer les organismes vivants actuels et fossiles dans une classification prédictive, pour pouvoir les identifier et prédire leurs propriétés biologiques, constitue l’objectif de la systématique (Tillier et al., 2000)*

Si la synonymie entre les termes *taxonomy* et *systematics* est implicite dans les définitions précédentes, il arrive que certains auteurs la mentionnent de manière explicite :

- (64) *I was therefore somewhat surprised recently to find out that I was attaching to the word in my own thoughts certain meanings which it did not hold for some other taxonomists to whom I talked. So I desire to make it clear that I am now using the terms taxonomy, classification, and systematics as absolutely synonymous, since they all inherently refer to the science of arrangement or classification (Blackwelder, 1940)*
- (65) *Taxonomy, or systematics, the first aspect of plant science to be studied, has consisted traditionnally of exploring the world for new kinds*

of plants, cultivating them in the garden, and studying them in the herbarium and library (Thimann et al., 1966)

- (66) According to Becklund, et al. (1969), systematics, in its broadest definition, is concerned with arranging the forms of the organic world into an observable scheme of relationships. The term "taxonomy" may be employed as a synonym for "systematics," but more often taxonomy describes the actual technical aspects of naming and classifying organisms. Both terms are useful and can be used interchangeably (Denmark, 1971)
- (67) Some authorities use the term 'Systematic biology' to describe this area of research [taxonomy]. For the present purposes, the two terms may be taken as interchangeable (Krebs, 1992)
- (68) taxonomy and systematics, the naming and classification of living thing (Thiele et Yeates, 2002)

En revanche, d'autres auteurs font une distinction nette entre les deux :

- (69) *Systematic Zoology, as I perceive it, consists of three main fields, taxonomy, nomenclature and phylogeny. These fields are constantly playing upon each other and being played upon by other fields of zoology and by other sciences* (Metcalf, 1937, p. 514)
- (70) *systematists (also known as taxonomists, there is a subtle difference) unravel the diversity of life by establishing classifications based on evolutionary relatedness* (Lane, 1992)
- (71) *Systematics discovers biological history : understanding the origins and relationships among extinct and extant life forms (O'Hara, 1988) and taxonomy is the science of naming and classifying these organisms (de Queiroz and Gauthier, 1994)* (Cotterill, 1995)
- (72) *If we make a distinction between taxonomy as the practice of describing and classifying organisms, and systematics as the basic study of the diversity of organisms and all their relationship* (Coomans, 2000)
- (73) *Taxonomy provides the bricks, and systematics the plan, with which the house of the biological sciences is built* (May, 2004)

Annexe E

Pas à pas de l'historiographe

Dans HistCite, deux paramètres peuvent être choisis :

- Les paramètres graphiques, tels que la taille de la police, des nœuds, etc ;
- Les paramètres de construction : quels vont être le nombre d'articles considérés ? Sur quels critères ? On peut choisir de représenter les liens
 - entre les N articles ayant le nombre de citations le plus important au sein du corpus (paramètres « LCS » et « value »)
 - entre les N articles ayant le nombre de citations le plus important dans tout le *WoS* (paramètres « GCS » et « value »)
 - entre tous les articles ayant au minimum X citations au sein du corpus (paramètres « LCS » et « count »)
 - entre tous les articles ayant au minimum X citations au sein du *WoS* (paramètres « GCS » et « count »)

Ici j'ai choisi de visualiser l'ensemble des liens de citations à l'intérieur du corpus, pour tous les articles ; donc les paramètres « LCS » et « count = 515 »). L'aperçu graphique donné par HistCite n'est pas facilement interprétable, à moins de réduire le nombre d'articles comme il a été fait figure 6.2 ; mais ce n'était pas le but ici puisque je voulais visualiser l'ensemble des liens de citation. J'ai donc choisi d'exporter les données du réseau pour le visualiser avec l'outil de cartographie Gephi.

Le format d'export le plus complet, « Pajek2 », ne permet pas de conserver l'association entre les nœuds du graphe et les informations que l'on possède sur les documents (titre, etc.). Or, étudier qualitativement le graphe sans avoir accès aux informations concernant les nœuds est impossible. Comme j'avais, d'une part, les informations contenues par HistCite, et d'autre part, la liste des nœuds lus par Gephi, j'ai créé une base de données relationnelles sous Access pour reconstruire le lien entre ces deux fichiers. J'y ai également ajouté des informations supplémentaires : extraction des taxons étudiés dans l'article, des pays des auteurs, et rattachement de

l'article à une spécialité : phylogénie, autre travail sur des spécimens de collection, données de distribution, ou autre discipline.

1. export de la liste tabulée des articles et des données HistCite au format .CSV (fichier 1)
2. export de l'historiographe au format Pajek2
3. import de l'historiographe sous Gephi
4. export des nœuds sous format CSV (fichier 2)
5. conversion des fichiers CSV de HisteCite et Gephi en TAB avec CSV2TAB pour que l'import se fasse correctement dans Access
6. import du fichier 1 dans une table Access
7. import du fichier 2 dans une table Access
8. création des champs URL, slice, taxon, coverage, country
9. ajout d'une table Countries et d'une table Taxa
10. remplissage des colonnes à l'aide de fonctions VBA et requêtes.
11. export du fichier complété en CSV
12. import dans Gephi pour compléter les nœuds.

Annexe F

Description et période de validité des 43 termes de premier niveau du *subject thesaurus* du *ZR*

MOT-CLÉ	VALIDITÉ	DESCRIPTION
<i>Ecology</i>	1864- <i>now</i>	Interrelationships between animals and their environment
<i>Habitat</i>	1864- <i>now</i>	Site and particular type of local environment occupied by an organism ; restricted to items where the habitat itself is of significance
<i>Behaviour</i>	1864- <i>now</i>	Any observable action or response of an animal ; includes general ethology
<i>Systematics</i>	1864- <i>now</i>	-
<i>Abiotic factors</i>	1864- <i>now</i>	External/environmental/administered influencing factors present in the wild or captive environment of an organism
<i>Nutrition</i>	1864- <i>now</i>	All aspects of nourishment, includes any mechanisms involved in the procurement, ingestion and digestion of food to provide energy for maintenance of life
<i>Evolution</i>	1864- <i>now</i>	Cumulative change in the characteristics of organisms or populations during successive generations related by descent
<i>Reproduction</i>	1864- <i>now</i>	-

... à suivre ...

TABLEAU F.1 – suite

MOT-CLÉ	VALIDITÉ	DESCRIPTION
<i>Biochemistry</i>	1864- <i>now</i>	Chemistry of the substances and metabolic processes of animals ; includes chemical structure, biosynthesis, metabolic pathways, etc
<i>Parasites diseases and disorders</i>	1864- <i>now</i>	All forms of pathogenic conditions, malfunctions and injuries. For non parasitic associations use ASSOCIATIONS
<i>Genetics</i>	1864- <i>now</i>	Origin and inheritance of observable characteristics
<i>Techniques</i>	1864- <i>now</i>	Techniques, methods, procedures or apparatus used in the study of zoology ; used only for items where the technique itself is being evaluated/discussed. For conservation techniques use CONSERVATION MEASURES
<i>Animals and man</i>	1864- <i>now</i>	Man's impact on and use of animals. For zoogeographical introduction by man use INTRODUCTION
<i>Conservation</i>	1864- <i>now</i>	Preservation of natural resources
<i>Life cycle and development</i>	1864- <i>now</i>	-
<i>Documentation</i>	1864- <i>now</i>	Various types and forms of information-carrying materials related to zoology, where the format is of primary significance ; not equivalent to document type of source item
<i>Zoogeography</i>	1864- <i>now</i>	Discussion of patterns, trends and means of animal dispersal and distribution. For distribution records use LAND ZONES or MARINE ZONES
<i>Biometrics</i>	1864- <i>now</i>	Measurements of the animal itself
<i>General morphology</i>	1864- <i>now</i>	Not Myxozoa, Protozoa. Not Protozoa. Overall external and internal anatomy of the animal and its appendages ; includes morphology of all life cycle stages. For protozoan morphology use CYTOLOGY or narrower terms. For morphology of specific organ systems use term for appropriate organ in other hierarchies

... à suivre ...

TABLEAU F.1 – suite

MOT-CLÉ	VALIDITÉ	DESCRIPTION
<i>Skeletal and supporting structures</i>	1864- <i>now</i>	Framework of various compositions used for support, protection and muscle attachment. For internal cytoskeleton use CYTOPLASMIC INCLUSIONS and/or MICROFILAMENTS AND MICROTUBULES
<i>Feeding</i>	1864-1991	General aspects of food procurement; includes food sources and feeding methods for 1978-1991. 1992- <i>now</i> : indexed under NUTRITION or narrower terms
<i>Whole animal physiology</i>	1864- <i>now</i>	Whole animal only. For physiology of specific tissues/organs use appropriate morphological term from other hierarchies
<i>Pollution</i>	1864- <i>now</i>	-
<i>Associations</i>	1864- <i>now</i>	Relationships between organisms of different species, with one or more parties gaining benefit from the relationship. For kleptoparasitism use FOOD ROBBING. For parasitic associations use PARASITES DISEASES AND DISORDERS
<i>Sensory reception</i>	1864- <i>now</i>	Structures and processes allowing animals to perceive/receive external and internal stimuli
<i>Nervous system</i>	1864- <i>now</i>	Not Protozoa. Not Myxozoa, Protozoa. Structures and processes involved in the coordination of animal activities through the transmission of nervous impulses. For protozoan neural morphology use ORGANELLES or narrower terms. For neuroendocrinology use ENDOCRINE SYSTEM and/or NEUROENDOCRINE SYSTEM
<i>Digestive system</i>	1864- <i>now</i>	Not Protozoa. Not Myxozoa, Protozoa. Digestive tract and its associated glands. For protozoan digestive morphology use ORGANELLES. For hepatopancreas use DIGESTIVE GLAND
<i>Endocrinology</i>	1864-1977 1992- <i>now</i>	For 1978-1991 indexed under ENDOCRINE SYSTEM. Structure and function of hormone producing organs, glands and cells and their secretions for 1992- <i>now</i>
<i>Colour and luminescence</i>	1864-1977 1992- <i>now</i>	For 1978-1991 indexed under COLOUR and/or LUMINESCENCE

... à suivre ...

TABLEAU F.1 – suite

MOT-CLÉ	VALIDITÉ	DESCRIPTION
<i>Integument</i>	1864- <i>now</i>	Not Protozoa. Not Myxozoa, Protozoa. Outer covering or layers of an organism
<i>Circulatory system</i>	1864- <i>now</i>	Not Protozoa. Not Myxozoa, Protozoa. Structural and functional aspects of circulatory and lymphatic systems, circulatory fluids and the mechanisms which circulate them throughout the body
<i>Immunology and repair mechanisms</i>	1864- <i>now</i>	Response of an animal to antigens, foreign bodies, substances, infectious agents and/or physical damage. For biochemistry of antigens use ANTI-GENS
<i>Locomotion</i>	1864- <i>now</i>	Mechanisms by which progressive motion of an animal is achieved; includes general items on motility. For locomotory activity period use ACTIVITY PATTERNS
<i>Cytology</i>	1864- <i>now</i>	Subcellular structures; includes morphology of protozoans
<i>Respiration</i>	1864-1977 1992- <i>now</i>	For 1978-1991 indexed under RESPIRATORY SYSTEM. Structures and functions which enable gaseous exchange to occur between an organism and its environment for 1992- <i>now</i>
<i>Musculature</i>	1864- <i>now</i>	Not Myxozoa, Protozoa. Not Protozoa. General/comprehensive studies involving the system or arrangement of body muscles; includes structures and functions associated with the contractile process and production of movement. For protozoan contractile processes use CELL CONTRACTION. For heart musculature use MYOCARDIUM and/or HEART. For neuromuscular junctions use SYNAPSES
<i>Biophysics</i>	1864- <i>now</i>	Physical properties of animals and their tissues. For fluorescence use LUMINESCENCE. For iridescence use STRUCTURAL COLOURS
<i>Excretion</i>	1864-1977 1992- <i>now</i>	For 1978-1991 indexed under EXCRETORY SYSTEM. Structural and functional aspects of elimination of waste products resulting from cellular metabolism for 1992- <i>now</i>

... à suivre ...

ANNEXE F : Description et période de validité des 43 termes de premier niveau du *subject thesaurus* du ZR

TABLEAU F.1 – suite

MOT-CLÉ	VALIDITÉ	DESCRIPTION
<i>Histology</i>	1864- <i>now</i>	Not Myxozoa, Protozoa. Not Protozoa. General studies of particular types of tissues. For histology of specific structures use structural terms from other hierarchies
<i>Attachment</i>	1864-1991	For 1992- <i>now</i> indexed under appropriate term for aspect of biology discussed
<i>Economics</i>	1864-1991	For 1992- <i>now</i> indexed under ANIMALS AND MAN or narrower terms. New/unusual or historical/conservational aspects only for 1978-1991
<i>Theoretical zoology</i>	1864-1978	For 1980- indexed under appropriate term for aspect of biology discussed
<i>Biology</i>	1864-1977	-

Annexe G

Thesaurus correspondant au terme *Systematics* du *ZR*

Systematics
 Systematic relationships
 Numericlature
 Revision
 Nomenclature
 New taxa
 Phylum nov
 Subphylum nov
 Superclass nov
 Class nov
 Subclass nov
 Infraclass nov
 Cohort nov
 Division nov
 Subdivision nov
 Section nov
 Subsection nov
 Superorder nov
 Order nov
 Suborder nov
 Infraorder nov
 Superfam nov
 Fam nov
 Subfam nov
 Supertribe nov
 Tribe nov
 Subtribe nov
 Infratribe nov
 Gen nov
 Subgen nov
 New species group
 Sp nov

Subsp nov
 Taxon nov
 New aberration
 New form
 New morph
 New race
 New variety
 New PhyloCode name
 Type genus
 Designated type genus for
 Designated type genus for new infratribe
 Designated type genus for new subtribe
 Designated type genus for new tribe
 Designated type genus for new supertribe
 Designated type genus for new subfamily
 Designated type genus for new family
 Designated type genus for new superfamily
 Type locality
 Type horizon+
 Type host
 Type species
 Designated type species for
 Designated type species for new subgenus
 Designated type species for new genus
 Type material
 Allotype
 Cotype
 Hapantotype
 Holotype
 Lectotype
 Lectotype designated
 Neotype
 Neotype designated
 Paralectotype
 Paratype
 Syntype
 Topotype
 Diagnosis
 Diagnostic characters
 Diagnostic characters of subordinate taxa
 Emended diagnosis
 Description
 First time described
 Redescription
 First time figured
 Combination
 Comb nov
 Combination revalidated
 Homonymy
 Synonymy

Junior synonym
Senior synonym
Syn nov
New synonym
Removal from synonymy
Spelling
Emendation
Available name
Invalid name
Nomen conservandum
Nomen correctum
Nomen dubium
Nomen imperfectum
Nomen inviolatum
Nom nov
Nomen perfectum
Nomen protectum
Nomen translatum
Nomen vanum
Preoccupied name replaced by
Replaced by
Replacement name
Valid name
Unavailable name
Nomen negatum
Nomen nudum
Nomen nullum
Nomen oblitum
Nomen vetitum
Lapsus calami
Authorship
Publication date
Etymology
Priority
Precedence
International Code of Zoological Nomenclature
Proposal to Commission
Comments on proposal to Commission
Proposed conservation
Proposed emendation
Proposed invalidation
Proposed suppression
Proposed validation
Action by Commission
Placed on Official Index
Placed on Official List
Recommendation by Commission
Refusal by Commission
Use of plenary powers
PhyloCode

Taxonomy

Numerical taxonomy

Classification

Included taxa

Theoretical taxonomy

Species inquirenda

Taxonomic unit

Phylum

Subphylum

Superclass

Class

Subclass

Infraclass

Cohort

Division

Subdivision

Section

Subsection

Superorder

Order

Suborder

Infraorder

Superfamily

Family

Subfamily

Supertribe

Tribe

Subtribe

Genus

Subgenus

Species

Subspecies

Aberration

Group

Form

Morph

Race

Variety

Taxonomic position

Referred to

Referred with status downgrading to

Referred with status upgrading to

Transferred from

Transferred with status downgrading from

Transferred with status upgrading from

Status

Stat nov

Stat rev

Raised from

Raised to

ANNEXE G : Thesaurus correspondant au terme *Systematics* du ZR

Key	Downgraded from Downgraded to
	Key to classes
	Key to subclasses
	Key to orders
	Key to suborders
	Key to superfamilies
	Key to families
	Key to subfamilies
	Key to tribes
	Key to subtribes
	Key to genera
	Key to subgenera
	Key to species groups
	Key to species
	Key to subspecies

Annexe H

Stratégie de recherche bibliographique sur le thème de la santé de la taxonomie

Cette annexe détaille la stratégie de recherche suivie pour élaborer un corpus d'articles traitant de l'état des lieux ou de la santé de la taxonomie sur une période de temps allant de la fin du XIX^e siècle à aujourd'hui.

Pour les articles et livres scientifiques, la recherche a été réalisée dans

- les grandes bases de données bibliographiques (*Web of Science*, Scopus, *Zoological Record*),
- les plateformes en ligne d'éditeurs (JSTOR),
- les catalogues de livres numérisés (Google Books, Biodiversity Heritage Library, Internet Archive, Hathi Trust Digital Library, Gallica, Europeana).

Pour les différents rapports et articles grand public, la recherche a été réalisée sur

- les sites web d'institutions particulièrement impliquées dans le sujet : GBIF, EDIT
- le catalogue de la bibliothèque centrale du MNHN
- le moteur de recherche généraliste Google.

Pour toutes ces recherches en ligne, différentes stratégies successives de requêtes ont été mises en œuvre. Une requête simple, à l'aide de mots-clés uniques, était réalisée dans le cas des sources anciennes (livres numérisés, articles du XIX^e siècle) : « taxinomie », « taxonomie » ; « taxonomy », descriptive », etc. Les résultats étaient ensuite parcourus et les sources d'intérêt retenues.

Différentes requêtes, plus complexes, étaient formulées lorsque la requête simple ren-

voyait trop de documents. Ces requêtes multiples, à géométrie variable, sont trop nombreuses pour pouvoir être répertoriées ici. En voici néanmoins deux exemples :

(crisis OR problem* OR save OR danger* OR progress) AND (taxonom*)
(taxonomists OR taxonomy) AND (lack OR need)

Une recherche thématique sur fiches cartonnées dans une grande bibliothèque étrangère (bibliothèque de l'Institut de Zoologie de Saint-Petersbourg) a également permis de compléter cette recherche numérique par mots-clés.

Une recherche itérative dans la bibliographie des différents articles trouvés a été réalisée lorsque cela était possible, i.e. pour les articles mentionnant des références.

L'inscription à deux listes de diffusion (Taxacom et The Systematics Discussion Group at UC Berkeley) m'a également permis de mettre en place une veille sur ces thématiques.

Pour finir, j'ai également effectué en mars 2011 une demande d'articles en langues autres que anglais ou français sur la liste Taxacom¹. La liste Taxacom était à ce moment-là constituée de 1 950 membres appartenant de manière identifiable aux pays suivants : Argentina, Australia, Austria, Belgium, Botswana, Brazil, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Egypt, Estonia, Ethiopia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, India, Iran, Ireland, Italy, Japan, Korea, Kuwait, Latvia, Lithuania, Malaysia, Malta, Mexico, Micronesia, Namibia, Netherlands, New Zealand, Nicaragua, Norway, Oman, Peru, Portugal, Romania, Russian Federation, Singapore, Slovakia, Slovenia, South Africa, Spain, Sweden, Switzerland, Taiwan, Tunisia, Turkey, United Kingdom, United States of America, Uruguay, Zambia, Zimbabwe. Les pays d'appartenance des membres ont été identifiés grâce aux suffixes de leurs E-mails². Parmi les 1 912 emails accessibles, 42 % désignaient un nom de pays. J'ai obtenu des réponses enthousiastes, mais très peu nombreuses (au nombre de 12).

Je n'ai pas abordé le sujet des sociétés savantes, actuellement au nombre de 17 000. Les introductions de congrès annuels de sociétés savantes sont toutefois un matériel qu'il serait très intéressant d'analyser dans une perspective d'histoire comparée.

1. <http://mailman.nhm.ku.edu/pipermail/taxacom/2011-March/071824.html>.

2. <http://www.cip.com.au/service/slist.html>

Bibliographie

- ACADÉMIE DES SCIENCES. 2011. *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs*. Académie des sciences. URL <http://www.academie-sciences.fr/activite/rapport/avis170111.pdf>. 7
- ADVISORY BOARD FOR THE RESEARCH COUNCILS. 1979. *Taxonomy in Britain*. Stationery Office Books. xiii, 119
- ANONYMOUS. 1946a. Genetics and taxonomy in India. *Nature* 157(3980) : 1. 110, 165
- ANONYMOUS. 1946b. Importance of taxonomy. *Nature* 158(4003) : 2. 47, 111, 125, 165
- ANONYMOUS. 1992. Global Biodiversity - Status of the Earth's living resources. *Rapport technique*, World Conservation Monitoring Centre. xi, xii, 104
- ANONYMOUS. 1995. *Global Biodiversity Assessment*. UNEP. 166
- ANONYMOUS. 2002. Genomics and taxonomy for all. *Nature* 417(6889) : 573–573. 114, 125
- ANONYMOUS. 2008a. *Government response to the House of Lords Science and Technology Committee Inquiry into Systematic Biology Research and Taxonomy*. UK Department of Health. 168
- ANONYMOUS. 2008b. Proceedings of the national taxonomy forum ; Australian Museum, Sydney ; 3–4 October 2007. In HALL, K., [éd.], *National taxonomy forum*, 64. Federation of Australian Scientific and Technological Societies (FASTS) / Australian Biological Resources Study (ABRS), Australian Museum, Sydney. 85, 168
- AUSTRALIAN GOVERNMENT DEPARTMENT OF THE ENVIRONMENT WATER HERITAGE AND THE ARTS. 2006. Survey of Australian Taxonomic Capacity. xiii, 104
- BAETHGE, C. 2008. Publish Together or Perish. The Increasing Number of Authors per Article in Academic Journals Is the Consequence of a Changing Scientific

- Culture. Some Researchers Define Authorship Quite Loosely. *Deutsches Ärzteblatt International* 105(20) : 380–383. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2696880/?tool=pubmed>. 89
- BALCELLS, E. 1963. Sugerencias y aspectos de labor zoologica en Espana. *Rapport technique*, Consejo Superior de Investigaciones Científicas. 119
- BALL, C. R. 1946. Why Is Taxonomy Ill-supported? *Science* 103(2685) : 2. 110, 111
- BARRATT, B. I. P., J. G. B. DERRAIK, C. G. RUFAUT, A. J. GOODMAN ET K. J. M. DICKINSON. 2003. Morphospecies as a substitute for Coleoptera species identification, and the value of experience in improving accuracy. *Journal of the Royal Society of New Zealand* 33(2) : 583–590. xii
- BATEMAN, R. 2008. When diagnosis is easier than cure. *Research Fortnight* 16–16. 51, 168
- BEETON, B. 2008. Taxonomy in peril. 168
- BENDER-DEMOLL, S. ET D. A. MCFARLAND. 2006. The Art and Science of Dynamic Network Visualization. *Journal of Social Structure* 7(2). 61
- BERGSTROM, C. T., J. D. WEST ET M. A. WISEMAN. 2008. The Eigenfactor (TM) Metrics. *Journal of Neuroscience* 28(45) : 11433–11434. 7
- BERNAL, J. D. 1939. *The Social Function of Science*. Routledge Kegan Paul Ltd., London. 52
- BERT, P. 1885. *Premières notions de zoologie (4^e ed.* G. Masson, Paris. URL <http://gallica.bnf.fr/ark:/12148/bpt6k29377z/>. 213
- BIBLIOTHÈQUE DE L'UNIVERSITÉ D'OTTAWA. 2011. Bases de données (A-Z) - Bibliothèque uOttawa [En ligne] Accédé le 24 mai 2011. URL <http://www.biblio.uottawa.ca/content-page.php?g=fr&s=mrt&c=src-bas&linxID=19521>. 17
- BIBLIOTHÈQUE NATIONALE DE FRANCE. 1996. Tous les savoirs du monde : en savoir plus : les métaphores du savoir. URL <http://classes.bnf.fr/dossism/sp-metap.htm>. 52
- BIOSIS. 1998. Web Access Available for BIOSIS Products From SilverPlatter. [En ligne] Accédé le 24 mai 2011. URL http://web.archive.org/web/20030427122935/http://www.biosis.org/news_events/pr_archive/1998/norwood.html. 18
- BIOSIS. 2000. BIOSIS and CSA Release Zoological Record Plus. [En ligne] Accédé le 24 mai 2011. URL http://web.archive.org/web/20030427125423/http://www.biosis.org/news_events/pr_archive/2000/zrplus_0700.html. 18

BIBLIOGRAPHIE

- BIOSIS. 2003. Zoological Record Increases Functionality. [En ligne] Accédé le 24 mai 2011. URL http://web.archive.org/web/20030604215510/http://biosis.org/news_events/press/2003/030529.html. 18
- BISBY, F. A., J. SHIMURA, M. RUGGIERO, J. EDWARDS ET C. HAEUSER. 2002. Taxonomy, at the click of a mouse. *Nature* 418 : 367–367. xii, 114
- BLACKMORE, S. 2002. Biodiversity Update - Progress in Taxonomy. *Science* 298(5592) : 365–365. 114
- BLACKWELDER, R. E. 1940. Some Aspects of Modern Taxonomy. *Journal of the New York Entomological Society* 48(3) : 245–257. 48, 169
- BLACKWELDER, R. E. 1959. The Present Status of Systematic Zoology. *Systematic Zoology* 8(2) : 69–75. 119, 165
- BOERO, F. 2010. The Study of Species in the Era of Biodiversity : A Tale of Stupidity. *Diversity* 2(1) : 115–126. 168
- BOLLEN, J., H. VAN DE SOMPEL, A. HAGBERG, L. BETTENCOURT, R. CHUTE, M. A. RODRIGUEZ ET L. BALAKIREVA. 2009. Clickstream Data Yields High-Resolution Maps of Science. *PLoS ONE* 4(3) : e4803. URL <http://dx.doi.org/10.1371/journal.pone.0004803>. 53
- BÖRNER, K. 2010. *Atlas of science - Visualizing what we know*. The MIT Press. 53
- BÖRNER, K., C. CHEN ET K. W. BOYACK. 2003. Visualizing Knowledge Domains. *Annual Review of Information Science & Technology* 37 : 58. URL <http://nwb.cns.iu.edu/papers/arist02.pdf>. 59
- BOSMAN, J., I. VAN MOURIK, M. RASCH, E. SIEVERTS ET H. VERHOEFF. 2006. *Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar*. Utrecht University Library. 73
- BOUCHET, P. ET C. M. DUARTE. 2006. The magnitude of marine biodiversity. In *The Exploration of Marine Biodiversity - Scientific and Technological Challenges*. Fundación BBVA. xiii
- BOUCHET, P., H. LE GUYADER ET O. PASCAL. 2009. The SANTO 2006 Global Biodiversity Survey : an attempt to reconcile the pace of taxonomy and conservation. *Zoosystema* 27(5) : 401–406. 126
- BOURNE, W. 1972. Taxonomy in distress. *Nature* 239(5367) : 1. 112
- BRADFORD, S. C. 1934. Sources of Information on Specific Subjects. *Engineering : an Illustrated Weekly Journal* 137 : 85–86. 5

- BRAUN, T. ET S. ZSINDELY. 1985. Growth of scientific literature and the Barnaby Rich effect. *Scientometrics* 7(3) : 529–530. 5
- BREMER, K., B. BREMER, P. O. KARIS ET M. KALLERSJO. 1990. Time for change in taxonomy. *Nature* 343(6255) : 202–202. 113, 125, 166
- BRIDSON, G. 1968. The Zoological Record - A Centenary Appraisal. *Journal of the Society for the Bibliography of Natural History* 5(1) : 23–34. 18
- BROWN, B. N. 2007. A comparative analysis of ecology literature databases. [En ligne] Accédé le 24 mai 2011. URL http://units.sla.org/division/dbio/events/conf_past/Denver/brownpaper.pdf. 17
- BRUES, C. T. 1939. Some Adaptive Responses of Taxonomy to a Changing Environment. *Journal of the New York Entomological Society* 47(2) : 145–154. 118, 213
- BRUMMITT, R. K. 1997. Taxonomy versus Cladonomy, a Fundamental Controversy in Biological Systematics. *Taxon* 46(4) : 723–734. 50
- CALLON, M., J.-P. COURTIAL ET H. PENAN. 1993. *La scientométrie*. Presses universitaires de France, Paris. 6, 11, 13, 14, 15
- CALLON, M., J. LAW ET A. RIP. 1986. *Mapping the Dynamics of Science and Technology : Sociology of Science in the Real World*. MacMillan, London. 52
- DE CANDOLLE, A. 1873. *Histoire des sciences et des savants depuis deux siècles, suivie d'autres études sur des sujets scientifiques, en particulier sur la sélection dans l'espèce humaine*. 1^{re} édition. H. Georg, Genève. 4, 213
- DE CANDOLLE, A. P. 1813. *Théorie élémentaire de la botanique : ou, Exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux*. Paris - Deterville. 45
- DE CARVALHO, M., F. BOCKMANN, D. AMORIM ET C. BRANDÃO. 2008. Systematics must Embrace Comparative Biology and Evolution, not Speed and Automation. *Evolutionary Biology* 35(2) : 150–157. 47, 167
- CATTELL, J. M. 1906. A statistical study of American men of science. *Science* 24 : 658–665. 4
- CENTRAL INTELLIGENCE AGENCY. 2008. The World Factbook 2008. 105
- CHALMERS, N., D. HAWKSWORTH, D. S. INGRAM, G. LONG, G. T. PRANCE, P. H. RAVEN ET L. E. SKOG. 1990. In defence of taxonomy. *Nature* 347(6290) : 224–224. 113

BIBLIOGRAPHIE

- CHEN, C. 2006. CiteSpace II : Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 57(3) : 359–377. 61
- CHISMAN, J. K. ET E. BREKKE. 1996. Comparing Coverage in 2 Indexes : Wildlife Review and Zoological Record. *Wildlife Society Bulletin* 24(1) : 149–152. 17
- CLIFFORD, H. T., R. W. ROGERS ET M. E. DETTMANN. 1990. Where now for taxonomy. *Nature* 346(6285) : 602–602. 113
- COBO, M. J., A. G. LÓPEZ-HERRERA, E. HERRERA-VIEDMA ET F. HERRERA. 2011. Science mapping software tools : Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology* n/a–n/a. 36, 61
- COCKERELL, T. 1946. Botanical taxonomy. *Science* 104(2693) : 1. 110, 111, 125
- COLE, F. ET N. EALES. 1917. The history of comparative anatomy. Part I : A statistical analysis of the literature. *Science Progress* 11 : 578–596. 5
- COLE, L. J. 1941. Each after His Kind. *Science* 93(2413) : 289–293. 169
- COLLIANDER, C. ET P. AHLGREN. 2011. Experimental comparison of first and second-order similarities in a scientometric context. *Scientometrics* 1–11. URL <http://www.springerlink.com/content/x147018216781x35/fulltext.pdf>. 59
- COOMANS, A. 2000. Nematode systematics : past, present and future. *Nematology* 2 : 3–7. 167, 170
- COOMANS, A. 2002. Present status and future of nematode systematics. *Nematology* 4 : 573–582. xiii
- COSTELLO, M. J., P. BOUCHET, C. S. EMBLOW ET A. LEGAKIS. 2006. European marine biodiversity inventory and taxonomic resources : state of the art and gaps in knowledge. *Marine Ecology Progress Series* 316 : 257–268. 125, 133, 167
- COTTERILL, F. P. D. 1995. Systematics, biological knowledge and environmental conservation. *Biodiversity and Conservation* 4 : 183–205. 170
- COURTIAL, J.-P. 1995. Construction des connaissances scientifiques, construction de soi et communication sociale. *Solaris* 2 : 201–217. 15
- COURTIAL, J.-P. 2006. L'Association pour la mesure des sciences et des techniques (Adest) et l'évaluation de la recherche en France. 4, 15
- CRISCI, J. 2006a. Making Taxonomy Visible. *Systematic Botany* 31(2) : 439–440. xii

- CRISCI, J. V. 2006b. One-Dimensional Systematist : Perils in a Time of Steady Progress. *Systematic botany* 31(1) : 217–221. 47, 167
- CUVIER, G. 1798. *Tableau élémentaire de l'histoire naturelle des animaux*. Baudouin, imprimeur, Paris. 126
- DAUGHERTY, C. H., A. CREE, J. M. HAY ET M. B. THOMPSON. 1990. Neglected taxonomy and continuing extinctions of Tuatara (*Sphenodon*). *Nature* 347(6289) : 177–179. 112, 113
- DAYRAT, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85(3) : 407–415. 77, 120
- DENMARK, H. 1971. Insect Identification : What Is It ? *The Florida Entomologist* 54(1) : 101–104. 170
- DIALOG. 2011. Dialog - Key Dates in the Company's History URL <http://www.dialog.com/about/keydates/>. 17
- DIDEROT, D. ET J. L. R. D'ALEMBERT. 1751. Discours préliminaire des éditeurs. *In Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*. les Sociétés Typographiques, Bern-Lauzanne. 52, 57, 207
- DOBZHANSKY, T. 1936. A Critique of the Species Concept in Biology. *Philosophy of Science* 2(3) : 344–355. 47, 165
- DUBOIS, A. 1997. An evolutionary biologist's view on the science of biology. *Alytes* 15(3) : 133–136. 50
- DUBOIS, A. 2003a. Authors of zoological publications and nomina are signatures, not persons. *Zootaxa* 1771 : 63–68. URL <http://www.mapress.com/zootaxa/2008/2/zt01771p068.pdf>. 34
- DUBOIS, A. 2003b. The relationships between taxonomy and conservation biology in the century of extinctions. *Comptes Rendus Biologies* 326(Supplement 1) : 9–21. 125, 133, 167
- DUBOIS, A. 2008a. Drôles d'espèces. Hybridation, perturbations de la méiose et spéciation dans le règne animal : quelques points délicats de terminologie, d'éidonomie et de nomenclature. *In* PRAT, D., A. RAYNAL-ROQUES ET A. ROGUENANT, [éds.], *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*. Belin, Paris. 50
- DUBOIS, A. 2008b. Handicap taxinomique et crise de la biodiversité : un nouveau paradigme pour la biologie au 21^e siècle. *In* PRAT, D., A. RAYNAL-ROQUES ET A. ROGUENANT, [éds.], *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*. Belin, Paris. 134

BIBLIOGRAPHIE

- DUBOIS, A. 2010. Zoological nomenclature in the century of extinctions : Priority vs. "usage". *Organisms, Diversity & Evolution* in press. 133
- DUBOIS, A. ET J. RAFFAËLLI. 2009. A new ergotaxonomy of the family Salamandridae Goldfuss, 1820 (Amphibia, Urodela). *Alytes* 26(1-4) : 1–85. 50
- DUNN, C. P. 2003. Keeping taxonomy based in morphology. *Trends in Ecology & Evolution* 18(6) : 270–271. xii
- VAN ECK, N. J. ET L. WALTMAN. 2009. VOSviewer : A Computer Program for Bibliometric Mapping. Slides. URL <http://www.vosviewer.com/publications/ModelingScience2009.pdf>. 60
- VAN ECK, N. J. ET L. WALTMAN. 2010. Software survey : VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2) : 523–538. 59, 61
- VAN ECK, N. J., L. WALTMAN, R. DEKKER ET J. VAN DEN BERG. 2010. A comparison of two techniques for bibliometric mapping : Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology* 61(12) : 2405–2416. 62
- EDWARDS, J. L., M. A. LANE ET E. S. NIELSEN. 2000. Interoperability of Biodiversity Databases : Biodiversity Information on Every Desktop. *Science* 289(5488) : 2312–2314. URL <http://www.sciencemag.org/content/289/5488/2312.short>. xv
- EDWARDS, S. R., G. M. DAVID ET L. NERLING. 1985. The systematics community. *Rapport technique*, Association of Systematics Collections c/o Museum of Natural History, University of Kansas, Lawrence, Kansas 66045. 119
- ELLINGHAM, H. J. T. 1948. Divisions of Natural Science and Technology. *In Report and Papers Submitted to the Royal Society Scientific Information Conference*. Burlington House, London. 58, 207
- ESPOSITO, M. 2011. The Impact Factor : Its Use, Misuse, and Significance. *International Journal of Prosthodontics* 24(1) : 85. 104
- EURONEWS. 2007. Thomson agrees to buy Reuters URL <http://www.euronews.net/2007/05/15/thomson-agrees-to-buy-reuters/>. 18
- EVANS, H. 1972. Modern taxonomy. *Science* 178(4066) : 2. 112, 166
- EVENHUIS, N. L. 2007. Helping Solve the “Other” Taxonomic Impediment : Completing the Eight Steps to Total Enlightenment and Taxonomic Nirvana. *Zootaxa* 1407 : 3–12. xii
- FALAGAS, M. E., V. D. KOURANOS, R. ARENCIBIA-JORGE ET D. E. KARAGEORGOPOULOS. 2008. Comparison of SCImago journal rank indicator with journal impact factor. *Faseb Journal* 22(8) : 2623–2628. 73

- FEINSINGER, P. 1996. Biodiversity knowledge in Chile : diagnosis and the first prescription. *Trends in Ecology & Evolution* 11(11) : 451–452. [xiii](#)
- FELLEGI, I. ET A. SUNTER. 1969. A Theory for Record Linkage. *Journal of the American Statistical Society* 64(328) : 1183–1210. [35](#)
- FISCHER, J.-L. ET R. REY. 1983. De l'origine et de l'usage des termes taxinomie-taxonomie. *Documents pour l'histoire du vocabulaire scientifique* V : 97–113. [46](#)
- FOSBERG, F. R. ET W. M. DIEHL. 1946. Present Status of Foreign Herbaria and Museums. *Science* 103(2670) : 2. [85](#), [110](#), [111](#)
- FROST, D. R. 2011. Amphibian Species of the World : an Online Reference. Version 5.5 (31 January, 2011). Electronic Database. URL <http://research.amnh.org/vz/herpetology/amphibia/>. [33](#)
- GALTON, F. 1874. *English men of science : their nature and nurture*. MacMillan, London. [4](#)
- GARFIELD, E. 1955. Citation indexes for science. A new dimension in documentation through association of ideas. *Science* 122(3159) : 108–111. [6](#)
- GEE, H. 1990a. Natural history museum - Taxonomy pays for bad image. *Nature* 345(6270) : 4–4. [112](#), [113](#)
- GEE, H. 1990b. One in six jobs to go. *Nature* 344(6269) : 805–805. [112](#)
- GEE, H. 1990c. Taxonomy - database for plant species. *Nature* 348(6299) : 275–275. [113](#)
- GEE, H. 1992. Peers slam peer review. *Nature* 355 : 488–488. [169](#)
- GEWIN, V. 2002. Taxonomy - All living things, online. *Nature* 418(6896) : 362–363. [xii](#), [114](#)
- GODFRAY, H. C. J. 2002a. Challenges for taxonomy. *Nature* 417(6884) : 17–19. [114](#), [167](#)
- GODFRAY, H. C. J. 2002b. How might more systematics be funded? *Antenna : Bulletin of the Royal Entomological Society* 26 : 11–17. [xii](#)
- GODFRAY, H. C. J. 2002c. Towards taxonomy's 'glorious revolution'. *Nature* 420(6915) : 461–461. [133](#)
- GODIN, B. 2005. *La science sous observation. Cent ans de mesure sur les scientifiques*. 1 édition. Les Presses de l'Université Laval, Sint-Nicolas. [4](#)

BIBLIOGRAPHIE

- GONZÁLEZ-OREJA, J. A. 2008. The Encyclopedia of Life vs. the Brochure of Life : Exploring the relationships between the extinction of species and the inventory of life on Earth. *Zootaxa* 1965 : 61–68. 133
- GONZÁLEZ-PEREIRA, B., V. GUERRERO-BOTE ET F. MOYA-ANEGÓN. 2010. A new approach to the metric of journals' scientific prestige : The SJR indicator. *Journal of Informetrics* . 7
- GRAHAM, M. W. R. d. V. 1972. Taxonomy in distress. *Nature* 238(5365) : 1. 112, 125
- GROSS, P. ET E. GROSS. 1927. College libraries and chemical education. *Science* 66 : 1229–1234. 5
- GUERRA-GARCIA, J. M., F. ESPINOSA ET J. C. GARCIA-GOMEZ. 2008. Trends in Taxonomy today : an overview about the main topics in Taxonomy. *Zoologica baetica* 19 : 15–49. 168
- HAAS, F. ET C. L. HÄUSER. 2006. How many taxonomists are there? [En ligne] Accédé le 24 mai 2011. URL http://www.gti-kontaktstelle.de/taxonomy_Lit.html. 50, 211
- HALL, H. M. ET F. E. CLEMENTS. 1923. *The phylogenetic method in taxonomy*. The Carnegie institution of Washington, Washington. 165
- HARLEY, R. M. 1990. In defence of taxonomy. *Nature* 347(6290) : 223–223. 113, 132, 166
- HATCH, M. H. 1948. The "Compleat" Taxonomist. *The Coleopterists Bulletin* 2(8) : 71–72. 120
- HERBERT, D. G., G. F. SMITH, M. HAMER ET C. H. SCHOLTZ. 2001. Taxonomy and systematics research in South Africa : vital research facing a crisis in capacity and resources. xiii
- HITCHCOCK, S. 1916. The Scope and Relations of Taxonomic Botany. *Science* 43(1106) : 331–342. 85, 118, 120, 169
- HITCHCOCK, S. 1928. The Need for Training Taxonomic Botanists. *Science* 67(1739) : 431–432. 118
- HOBERG, E. P., P. A. PILITT ET K. E. GALBREATH. 2009. Why Museums Matter : A Tale of Pinworms (Oxyuroidea : Heteroxynematidae) Among Pikas (*Ochotona princeps* and *O. collaris*) in the American West. *Journal of Parasitology* 95(2) : 490–501. 134

- HOLYŃSKI, R. B. 2008. God save us from friends or How to ensure taxonomy against passing away by dint of a successful operation? *Polish taxonomical monographs* 15 : 3–54. 125, 168
- HOUSE OF LORDS SCIENCE AND TECHNOLOGY COMMITTEE. 2008. Systematics and Taxonomy : Follow-up. 5th report. [En ligne] Accédé le 24 mai 2011. 386. URL <http://www.parliament.the-stationery-office.co.uk/pa/ld200708/ldselect/ldsctech/162/162.pdf>. xiii, 104, 168
- HU, J. 1997. Zoological Record and its indexing methods. *Chinese Journal of Zoology* 32(4) : 34–36. 25
- HUGHES, J. 2001. Characterization of Unique Serials Indexed in the Zoological Record. [En ligne] Accédé le 10 mars 2010. 30(Spring 2001). URL <http://www.library.ucsb.edu/ist1/01-spring/refereed.html>. 17
- HULCR, J. 2009. Some repentance would not hurt taxonomy either : a junior taxonomist's response to Quentin Wheeler. *Systematic Entomology* 34 : 199–201. 168
- ISELY, D. 1972. The Disappearance. *Taxon* 21(1) : 3–12. 51
- JACOMY, M., S. HEYMAN, T. VENTURINI ET M. BASTIAN. 2011. ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization. *in prep* URL http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf. 78
- JANZEN, D. H. 2004. Now is the time. *Philosophical Transactions of the Royal Society B* 359 : 731–732. 167
- JARZEMBOWSKI, E. A. 1990. In defence of taxonomy. *Nature* 347(6290) : 222–222. 113, 132
- JOPPA, L. N., D. L. ROBERTS ET S. L. PIMM. 2011. The population ecology and social behaviour of taxonomists. *Trends in Ecology & Evolution* 26(11) : 551 – 553. URL <http://www.sciencedirect.com/science/article/pii/S0169534711002084>. 103, 106
- JORDAN, D. S. 1888. The nomenclature of American Birds. *In Science sketches*. A.C. McClurg and company, Chicago. 47, 169
- KEMP, W. M. 1989. Presidential Address : Parasitology : A Degenerate Discipline, Populated by Degenerate Scientists, Studying Degenerate Organisms? *The Journal of Parasitology* 75(6) : 817–823. 51
- KLOPPER, R. R., G. F. SMITH ET A. C. CHIKUNI. 2002. The Global Taxonomy Initiative in Africa. *Taxon* 51(1) : 159–165. xiii, 47, 167
- KNAPP, S. 2007. Taxonomy as a team sport. *In The new taxonomy*. Wheeler, Q.D. 45, 167

BIBLIOGRAPHIE

- KNAPP, S., R. M. BATEMAN, N. R. CHALMERS, C. J. HUMPHRIES, P. S. RAINBOW, A. B. SMITH, P. D. TAYLOR, R. I. VANE-WRIGHT ET M. WILKINSON. 2002. Taxonomy needs evolution, not revolution - Some changes are clearly necessary, but science cannot be replaced by informatics. *Nature* 419(6907) : 559–559. [xii](#), [47](#), [85](#), [114](#), [167](#)
- KOSZTARAB, M. 1984. A Biological Survey of the United States. *Science* 223(4635) : 443. [125](#), [132](#)
- KREBS, J. R. 1992. *Evolution and Biodiversity : The New Taxonomy*. UK Natural Environment Research Council, London. [166](#), [170](#)
- KRELL, F.-T. 2002. Why impact factors don't work for taxonomy. *Nature* 415(6875) : 957–957. [114](#)
- LA SALLE, J., Q. WHEELER, P. JACKWAY, S. WINTERTON, D. HOBERN ET D. LOVELL. 2009. Accelerating taxonomic discovery through automated character extraction. *Zootaxa* 2217 : 43–55. [168](#)
- LANE, R. P. 1992. The 'new taxonomy' - does it require new taxonomists or a new understanding? *Bulletin of entomological research* 82 : 437–440. [170](#)
- LEMERCIER, C. ET C. ZALC. 2008. *Méthodes quantitatives pour l'historien*. La Découverte. [xv](#)
- LEYDESDORFF, L. 2001. Scientometrics and Science Studies. *In The Challenge of Scientometrics. The Development, Measurement, and Self-Organization of Scientific communications*. Universal Publishers. [15](#)
- LIU, X., L. ZHANG ET S. HONG. 2011. Global biodiversity research during 1900-2009 : a bibliometric analysis. *Biodiversity Conservation* 20 : 807–826. [104](#)
- LOEBL, I. ET R. A. B. LESCHEN. 2005. Demography of Coleopterists and Their Thoughts on DNA Barcoding and the Phylocode, with Commentary. *The Coleopterists Bulletin* 59(3) : 284–292. [xiii](#), [120](#)
- LOTKA, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12) : 317–323. [5](#)
- LULLE, R. 1515. *Arbor scientie / Raymundi Lullii et in cujus commendationes est hoc extemporanei Jodoci Badii Ascensii ad pium lectorem epigramma*. in domo Francisci Fradin impressosis, Lugdunum. [52](#), [55](#), [207](#)
- MACPHERSON, E. 1990. Stable taxonomy. *Nature* 345(6275) : 485–485. [113](#)
- MALLIG, N. 2010. A relational database for bibliometric analysis. *Journal of Informetrics* 4(4) : 564–580. URL <http://ideas.repec.org/p/zbw/fisidp/22.html>. [61](#)

- MARQUES, A. C. ET C. J. E. LAMAS. 2006. Sistemática zoológica no Brasil : estado da arte, expectativas e sugestões de ações futuras. *Papéis Avulsos de Zoologia* 46(3) : 139–174. xiii
- MATHER, P. 2004. Taxonomy in Australia today. xiii, 167
- MAY, R. M. 1990. Taxonomy as destiny. *Nature* 347(6289) : 129–130. 113
- MAY, R. M. 2004. Tomorrow's taxonomy : collecting new species in the field will remain the rate-limiting step. *Philosophical Transactions of the Royal Society B* 359(1444) : 733–734. 48, 170
- MAYR, E. 1989. Future challenges of Systematic Ornithology. *Journal of the Yamashina Institute for Ornithology* 21(2) : 154–164. 51, 166
- MAYR, E. ET P. D. ASHLOCK. 1991. *Principles of systematic zoology. Second edition.* McGraw-Hill, New-York. 50, 166
- MCCALLUM, M. L. ET J. MCCALLUM. 2006. Publication Trends of Natural History and Field Studies in Herpetology. *Herpetological Conservation and Biology* 1 : 62–67. 167
- MCCLAIN, C. 2011. The Mass Extinction of Scientists Who Study Species. [En ligne] Accédé le 18 février 2011. URL <http://www.wired.com/wiredscience/2011/01/extinction-of-taxonomists/>. xi
- MCCOOK, A. 2004. Thomson acquires Biosis - The Scientist - Magazine of the Life Sciences. 18
- MEDLINE. 2011. Number of Authors per MEDLINE®/PubMed®Citation. URL <http://www.nlm.nih.gov/bsd/authors1.html>. 89
- MERRIAM, C. H. 1893. Biology in Our Colleges : A Plea for a Broader and More Liberal Biology. *Science* 21(543) : 352–355. 51, 118
- METCALF, Z. 1937. How the Taxonomist Names the Animals. *The Scientific Monthly* 45(6) : 513–523. 48, 165, 170
- METCALF, Z. 1946. Taxonomy and the biologists. *Science* 104(2701) : 2. 85, 110, 111, 165
- MICHÁN, L. ET J. LLORENTE BOUSQUETS. 2009. Bibliometría de la sistemática biológica sobre América Latina durante el siglo XX en tres bases de datos mundiales. *Revista de Biología Tropical* 58(2) : 531–545. xiii
- MICHÁN, L., J. M. RUSSELL, A. SANCHEZ PEREYRA, A. LLORENS CRUSET ET C. LÓPEZ BELTRÁN. 2008. Análisis de la sistemática actual en Latinoamérica. *Interciencia* 33(10) : 754–761. xiii, 168

BIBLIOGRAPHIE

- MOYA ANEGÓN, F., V. H. SOLANA ET E. JIMÉNEZ-CONTRERAS. 2006. A connectionist and multivariate approach to science maps : the SOM, clustering and MDS applied to library and information science research. *J. Information Science* 32(1) : 63–77. 61
- O'HARA, R. J., D. R. MADDISON ET P. F. STEVENS. 1988. Crisis in Systematics. *Science* 241(4863) : 275–276. 169
- OLIVER, J. H. 1988. Crisis in Biosystematics of Arthropods. *Science* 240(4855) : 967. 85
- ONU. 2011. Transcription du chinois. URL http://portal.unesco.org/culture/fr/files/32315/11625533133chinese_fr.pdf/chinese_fr.pdf. 37
- OREN, A. ET E. STACKEBRANDT. 2002. Prokaryote taxonomy online : challenges ahead. *Nature* 419(6902) : 15–15. 114
- ORTNER, H. M. 2011. The impact factor and other performance measures - much used with little knowledge about. *International Journal of Refractory Metals and Hard Materials* 28(5) : 559–566. 104
- PALLINI, A., M. A. M. FADINI, M. VENZON, G. J. DE MORAES ET D. M. BARROS-BATTESTI. 2007. Demandas e perspectivas para a Acarologia no Brasil (Demands and perspectives to acarology in Brazil). *Neotropical Biology and Conservation* 2(3) : 169–175. 125
- PAQUIN, P., N. DUPERRE, K. COKENDOLPHER, James C. and White ET M. HEDIN. 2008. The fundamental importance of taxonomy in conservation biology : the case of the eyeless *Cicurina bandida* (Araneae : Dictynidae) of central Texas, including new synonyms and the description of the male of the species. *Invertebrate Systematics* 22(2) : 139–149. 126
- PARKER, J. 2005. Evaluating bibliographic database overlap for marine science literature using an ecological concept. [En ligne] Accédé le 9 mars 2010. 42(Spring 2005). URL <http://www.ist1.org/05-spring/refereed-2.html>. 17
- PEARSON, D. L., A. L. HAMILTON ET T. L. ERWIN. 2011. Recovery Plan for the Endangered Taxonomy Profession. *BioScience* 61(1) : 58–63. xi
- PENEV, L., M. SHARKEY, T. ERWIN, S. VAN NOORT, M. BUFFINGTON, K. SELTMANN, N. JOHNSON, M. TAYLOR, F. C. THOMPSON ET M. J. DALLWITZ. 2009. Data publication and dissemination of interactive keys under the open access model. *ZooKeys* 21 : 1–17. xii
- POLANCO, X. 1995. Aux sources de la scientométrie. In *Les sciences de l'information : bibliométrie, scientométrie, infométrie*, tome 2. Presses universitaires de Rennes, Rennes. 3, 4

- POLASZEK, A., M. ALONSO-ZARAZAGA, P. BOUCHET, D. J. BROTHERS, N. EVENHUIS, F.-T. KRELL, C. H. C. LYAL, A. MINELLI, R. L. PYLE, N. J. ROBINSON, F. C. THOMPSON ET J. VAN TOL. 2005. ZooBank : the open-access register for zoological taxonomy : Technical Discussion Paper. *Bulletin of Zoological Nomenclature* 62(4) : 210–220. xii
- PRATHAPAN, K. D., D. R. PRIYADARSANAN, T. C. NARENDRAN, C. A. VIRAKTAMATH, N. A. ARAVIND ET J. POORANI. 2008. Death sentence on taxonomy in India. *Current science* 94(2) : 170–171. 168
- PRITCHARD, A. 1969. Statistical Bibliography or Bibliometrics? *Journal of Documentation* 25(4) : 348–349. 5
- PROQUEST. 2010. Zoological Record Plus (ZR Plus). [En ligne] Accédé le 24 mai 2011. URL <http://www.csa.com/factsheets/zooclust-set-c.php>. 17
- DE QUEIROZ, K. ET J. GAUTHIER. 1992. Phylogenetic taxonomy. *Annual Review of Ecology and Systematics* 23 : 449–480. URL <http://www.faculty.biol.ttu.edu/strauss/Phylogenetics/Readings/deQueirozGauthier1992.pdf>. 166
- QUINT, B. 2003. BIOSIS Sold to Thomson Scientific. 18
- R DEVELOPMENT CORE TEAM. 2011. R : a language and environment for statistical computing. 28
- RACOVITZA, É. G. 1907-1908. *Archives de zoologie expérimentale et générale : histoire naturelle, morphologie, histologie, évolution des animaux...*, tome 7 de 4. Germer Baillière, Paris. URL <http://gallica.bnf.fr/ark:/12148/bpt6k5453901s/f15.image.r=taxonomistes.langFR>. 213
- RAFOLS, I., A. L. PORTER ET L. LEYDESDORFF. 2010. Science overlay maps : A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology* 61(9) : 1871–1887. URL <http://dx.doi.org/10.1002/asi.21368>. 53, 61
- RAVEN, P. H. 1972. Modern taxonomy. *Science* 178(4066) : 2. 133
- RAVEN, P. H. 2004. Taxonomy : where are we now? *Philosophical Transactions of the Royal Society B* 359(1444) : 729–730. 167
- RAVEN, P. H., B. BERLIN ET D. E. BREEDLOVE. 1971. The Origins of Taxonomy. *Science* 174 : 4. 112
- RICHARDSON, B. J. 1984. Identifying the Australian fauna : what remains to be done? *Search* 14 : 320–323. 166
- RIDE, W. D. L. ET A. MCCUSKER. 1978. Australian Biological Resources Study, 1973-78. *Rapport technique*, Australian Government Publishing Service. 119

BIBLIOGRAPHIE

- ROBINSON, B. L. 1901. Problems and Possibilities of Systematic Botany. *Science* 14(352) : 465–474. 125, 132
- RODMAN, J. E. ET J. H. CODY. 2003. The taxonomic impediment overcome : NSF's partnerships for enhancing expertise in taxonomy (PEET) as a model. *Systematic biology* 52(3) : 428–435. xii
- RODRIGUES, A. S. L., C. L. GRAY, B. J. CROWTER, R. M. EWERS, S. N. STUART, T. WHITTEN ET A. MANICA. 2010. A Global Assessment of Amphibian Taxonomic Effort and Expertise. *Bioscience* 60(10) : 798–806. xiii
- ROSSI, P. 1993. *Clavis universalis : arts de la mémoire, logique combinatoire et langue universelle de Lulle à Leibniz*. Jérôme Millon. 55
- ROSTAING, H. 1996. *La bibliométrie et ses techniques*. Marseille : Centre de recherche rétrospective, Toulouse : Sciences de la société. 5, 6, 7, 10, 11, 12, 207
- SALAZAR VALLEJO, S. L., N. E. GONZALEZ ET E. SCHWINDT. 2008. Taxonomia de invertebrados marinos : necesidades en latinoamérica (Marine invertebrate taxonomy : needs in Latin America). *Interciencia* 33(7) : 510–517. 168
- SARIYAR, M. ET A. BORG. 2010. The RecordLinkage Package : Detecting Errors in Data. *The R Journal* 2(2) : 61–67. 37
- SAVIGNY, C. d. 1587. *Tableaux accomplis de tous les arts libéraux , contenant... par singulière méthode de doctrine une générale et sommaire partition des dicts arts amassez et réduicts en ordre pour le soulagement et profit de la jeunesse, par M. Christofle de Savigny,...* J. et F. de Gourmont frères, Paris. 52, 56, 207
- SCIMAGO. 2007. SJR — SCImago Journal Country Rank. URL <http://www.scimagojr.com>. 60
- SCOBLE, M. J. 2004. Unitary or unified taxonomy? *Philosophical Transactions of the Royal Society B* 359(1444) : 699–710. xii
- SHANTZ, H. 1934. Botanical Research. *Science* 79(2051) : 351–355. 125
- SHAPIRO, F. R. 1992. Origins of Bibliometrics, Citation Indexing, and Citation Analysis : The Neglected Legal Literature. *Journal of the American Society for Information Science and Technology* 43(5) : 337–339. 5
- SHEAR, C. 1934. Mycology, Scientific and Otherwise. *Mycologia* 26(3) : 201–209. 51
- SHEAR, C. ET F. E. CLEMENTS. 1926. The Condition and Needs of Systematic Mycology. *Science* 63(1633) : 393–395. 118
- SHELDON, F. H., J. V. REMSEN ET R. T. BRUMFIELD. 2005. Academy of Natural Sciences : Job cuts. *Science* 307(5715) : 1560–1560. 167

- SHIMURA, J. 2003. *Global Taxonomy Initiative in Asia. Research Report from the National Institute for Environmental Studies*. National Institute for Environmental Studies. xiii
- SILBERMANN, G. 1835. *Énumération des entomologistes vivans; suivie de Notes sur les collections entomologiques des principaux musées d'histoire naturelle d'Europe, sur les Sociétés d'entomologie, sur les recueils périodiques consacrés à l'étude des insectes, et d'une table alphabétique des résidences des entomologistes* ["sic"], tome 1. Roret, Paris. 85
- SIMPSON, G. G. 1961. *Principles of Animal Taxonomy*. Columbia University Press. 47, 166
- SMALL, H. 1973. Co-citation in the scientific literature : A new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24 : 265–269. 52
- SMART, J. 1940. Entomological Systematics examined as a practical problem. In HUXLEY, J., [éd.], *The New Systematics*. Clarendon Press, Oxford. 169
- SMITH, G. F., M. BUYS, M. WALTERS, D. HERBERT ET M. HAMER. 2008. Taxonomic research in South Africa : the state of the discipline. *South African Journal of Science* 140(7-8) : 254–256. xiii, 85, 168
- DE SOLLA PRICE, D. J. 1963. *Little Science, big science*. Columbia University Press. 5
- DE SOLLA PRICE, D. J. 1972. *Science et supra-science*. Fayard, Paris. 5
- SPARTACO, G. ET A. GIOVANNI. 2007. The problem of subspecies and biased taxonomy in conservation lists : the case of mammals. *Folia Zoologica* 56(2) : 113–117. 126
- VAN DER SPOEL, S., A. PIERROT-BULTS ET R. VAN SOEST. 1972. Modern taxonomy. *Science* 178(4066) : 1. 112, 120, 166
- STEBBINS, G. L. 1969. The Significance of Hybridization for Plant Taxonomy and Evolution. *Taxon* 18(1) : 26–35. URL <http://www.jstor.org/stable/1218589>. 47, 166
- STEINS, J. L. 2001. Anthropology in Zoological Record. [En ligne] Accédé le 9 mars 2010. 16(1). URL <http://anssacrl.wordpress.com/publications/reviews/historical-reviews/zoological-record/>. 17, 19
- STEVENS, P. F. 1990. In defence of taxonomy. *Nature* 347(6290) : 222–223. 113
- STEYSKAL, G. C. 1965. Trend Curves of the Rate of Species Description in Zoology. *Science* 149(3686) : 880–882. 119, 132

BIBLIOGRAPHIE

- STIRTON, C. H., L. BOULOS, T. D. MACFARLANE, N. P. SINGH ET A. NICHOLAS. 1990. In defence of taxonomy. *Nature* 347(6290) : 223–224. 113, 125
- SUAREZ, A. V. ET N. D. TSUTSUI. 2004. The Value of Museum Collections for Research and Society. *BioScience* 54(1) : 66–74. 134
- SZWEDO, J. ET D. IWAN. 2008. To survive or to flourish ? Status and role of natural history museums in the biodiversity research. *Polish taxonomical monographs* 15 : 55–69. 134, 168
- TANCOIGNE, E. 2008. *Métataxonomie : l'état de la taxinomie zoologique à l'aube du 21^e siècle. Rapport de Master*. Muséum national d'Histoire Naturelle, Paris. xiv
- TANCOIGNE, E., C. BOLE, A. SIGOGNEAU ET A. DUBOIS. 2011. Insights from Zoo-taxa on potential trends in zoological taxonomic activity. *Frontiers in Zoology* 8(5). 106
- TANSLEY, A. G. 1939. Introduction to a symposium on the reciprocal relationship of ecology and taxonomy. *Journal of Ecology* 27 : 401–402. 118
- TAUTZ, D., P. ARCTANDER, A. MINELLI, R. H. THOMAS ET A. P. VOGLER. 2002. DNA points the way ahead of taxonomy - In assessing new approaches, it's time for DNA's unique contribution to take a central role. *Nature* 418(6897) : 479–479. 114, 125
- TAUTZ, D., P. ARCTANDER, A. MINELLI, R. H. THOMAS ET A. P. VOGLER. 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution* 18(2) : 70–74. 167
- THIELE, K. ET D. YEATES. 2002. Tension arises from duality at the heart of taxonomy - Names must both represent a volatile hypothesis and provide a key to lasting information. *Nature* 419(6905) : 337–337. xii, 48, 114, 167, 170
- THIMANN, K. V., R. BURRIS, L. CONSTANCE, K. ESAU, P. MANGELSDORF, A. RIKER, G. STEBBINS, W. C. STEERE ET P. R. STOUT. 1966. The Plant Sciences, Now and in the Coming Decade. 132, 170
- THOMAZ, P. G., R. S. ASSAD ET L. F. P. MOREIRA. 2011. Using the Impact Factor and H Index to Assess Researchers and Publications. *Arquivos Brasileiros de Cardiologia* 96(2) : 90–93. 104
- THOMSON REUTERS. 2008. Thomson completes acquisition of Reuters; Thomson Reuters shares begin trading today. [En ligne] Accédé le 24 mai 2011. URL http://www.thomsonreuters.com/content/press_room/corp/corp_news/212378. 19
- THOMSON REUTERS. 2009. Zoological Record Help - Advanced Search Field Tags. 18, 23, 211

- THOMSON REUTERS. 2011a. Company History - Thomson Reuters. [En ligne] Accédé le 24 mai 2011. URL http://thomsonreuters.com/about/company_history/#1890_-_1799. 19
- THOMSON REUTERS. 2011b. Zoological Records - Digital Archive - Science - Thomson Reuters. [En ligne] Accédé le 24 mai 2011. URL <http://science.thomsonreuters.com/support/faq/zr/>. 26
- THORNE, J. 2003. The Zoological Record and registration of new names in zoology. *The Linnean* 19(2) : 22–26. 18
- THORPE, W. 1940. Ecology and the future of Systematics. In HUXLEY, J., [éd.], *The New Systematics*. Clarendon Press, Oxford. 165
- TILLIER, S. ET P. DE WEVER. 2000. *Systématique : ordonner la diversité du vivant - Rapport sur la science et la technologie n°11*. Rapport sur la science et la technologie. Editions Tec & Doc, Paris. 166
- TILLIER, S., P. JANVIER ET A. DE RICQLÈS. 2000. Situation et définitions de la systématique. In TILLIER, P., Simon ; de Wever, [éd.], *Systématique : ordonner la diversité du vivant - Rapport sur la science et la technologie n°11*. Editions Tec & Doc, Paris. xiii, 169
- TOWNSEND, P. A. 2006. Taxonomy is important in conservation : a preliminary reassessment of Philippines species-level bird taxonomy. *Bird Conservation International* 16(2) : 155–173. 126
- TURRILL, W. B. 1958. Symposium on cyto-taxonomy. *Proceedings of the Linnean Society of London* 169(1-2) : 110–134. 213
- ULISES MOULINES, C. 2006. *La philosophie des sciences, fin XIXè / début XXIè siècle : l'invention d'une discipline*. Rue d'Ulm/Presses de l'Ecole normale supérieure, Paris. 3
- UNIVERSITY OF PITTSBURGH LIBRARY. 2007. Zoological Record Database Description. [En ligne] Accédé le 24 mai 2011. URL http://www.library.pitt.edu/articles/database_info/zool_rec.html. 17
- VALDECASAS, A. G., E. BELLO ET J. M. BECERRA. 1992. La Taxonomía en España. *Arbor : Ciencia, Pensamiento y Cultura* 561 : 77–102. xiii
- VELT, R. 2011. Cartographies des connaissances scientifiques. [En ligne. Accédé le 26 octobre 2011] URL <http://www.knowtex.com/blog/cartographier-sciences/>. 52
- VINKLER, P. 1988. An attempt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics* 13(5-6) : 239–259. 12, 13, 211

BIBLIOGRAPHIE

- WEED, C. M. 1894. Zoology in the High School. *The American Naturalist* 28(336) : 1003–1010. 120
- WERNER, Y. L. 2006. The case of impact factor versus taxonomy : a proposal. *Journal of Natural History* 40(21-22) : 1285–1286. URL <http://www.tandfonline.com/doi/abs/10.1080/00222930600903660>. 64
- WEST, J. G. ET B. J. CONN. 1990. In defence of taxonomy. *Nature* 347(6290) : 222–222. 113
- WHEELER, Q. D. 2007. Invertebrate systematics or spineless taxonomy? *Zootaxa* 1668 : 11–18. 47, 167
- WHEELER, Q. D. 2009. Revolutionary thoughts on taxonomy : declarations of independence and interdependence. *Zoologia* 26(1) : 1–4. xii
- WHEELER, Q. D., P. H. RAVEN ET E. O. WILSON. 2004. Taxonomy : impediment or expedient? *Science* 303 : 285–285. 85, 126
- WHITE, R. E. 1975. Trend Curves of the Rate of Species Description for Certain North American Coleoptera. *The Coleopterists Bulletin* 29(4) : 281–295. 119
- WIEGAND, K. M. 1936. Methods of Approach in Teaching Taxonomy. *Science* 83(2152) : 291–294. 85
- WIKIPEDIA. 2011. Thésaurus - Wikipédia URL <http://fr.wikipedia.org/wiki/Thesaurus>. 20
- WILEY, E. ET B. S. LIEBERMAN. 2011. *Phylogenetics : Theory and Practice of Phylogenetic Systematics*. 2d. édition. John Wiley and Sons Ltd, Chicester/Un. 168
- WILSON, E. 2003. The encyclopedia of life. *Trends in Ecology & Evolution* 18(2) : 77–80. URL <http://bugs.morphobank.org/articles/tree.pdf>. 47, 167
- WILSON, E. O. 1972. Modern taxonomy. *Science* 178(4066) : 1. 112, 166
- WILSON, E. O. 1985a. The biological diversity crisis : a challenge to science. *Issues in Science and Technology* 2(1) : 22. URL <http://www.nysm.nysed.gov/bri/pdf/biodiv.pdf>. 133
- WILSON, E. O. 1985b. Time to Revive Systematics. *Science* 230(4731) : 1227. 47, 169
- WILSON, E. O. 2004. Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society B* 359(1444) : 739–739. 47, 167

BIBLIOGRAPHIE

- WINKLER, W. E. 2006. *Overview of Record Linkage and Current Research Directions*. U.S. Census Bureau. 35
- WOODLAND, D. W. 2007. Are botanists becoming the dinosaurs of biology in the 21st century? *South African Journal of Botany* 73(3) : 343–346. 51
- YANCEY, R. 2005. Fifty years of citation indexing and analysis. [En ligne] Accédé le 8 juillet 2011. URL <http://science.thomsonreuters.com/news/2005-08/8289803/>. 6
- ZIPF, G. K. 1935. *The psycho-biology of language ; an introduction to dynamic philology*. Houghton Mifflin company, Boston. 5

Liste des figures

2.1	Distribution bibliométrique. D'après Rostaing, 1996, p.26.	10
2.2	Nombre de travaux par auteur pour tous les travaux portant sur les Amphibiens contenus dans le <i>Zoological Record</i>	11
2.3	Nombre d'articles par journal pour tous les travaux portant sur les Amphibiens contenus dans le <i>ZR</i>	12
3.1	Interface de recherche du <i>ZR</i> dans le portail Web of Knowledge de Thomson Reuters.	19
3.2	Exemple de référence issue du <i>Zoological Record</i> [v.4] et exportée au format balisé.	21
3.3	Informations contenues dans la division « Subject hierarchy » du thésaurus du <i>ZR</i> pour le mot-clé <i>taxonomy</i>	24
3.4	Temps de latence entre publication et informatisation dans le <i>ZR</i>	27
3.5	Nombre total d'articles dans le <i>ZR</i> de 1864 à 2010	28
3.6	Photographies de Pelobatidae.	34
3.7	Les différentes erreurs que l'on peut retrouver dans la dénomination des auteurs.	35
3.8	Répartition des paires d'auteurs en fonction de la mesure de similarité de jarowinkler	40
3.9	Répartition des prénoms en fonction de leur longueur	41
3.10	Augmentation du nombre de paires de noms de famille à comparer lorsque le seuil de similarité entre ces paires augmente	42
4.1	L' <i>Arbor scientie</i> de Lulle (1515)	55
4.2	<i>Encyclopedie</i> tirée des <i>Tableaux accomplis de tous les arts libéraux</i> de Savigny (1587)	56
4.3	<i>Système figuré des connaissances humaines</i> tiré du <i>Discours préliminaire des éditeurs</i> de Diderot et d'Alembert (1751). Source : ARTFL Encyclopédie Project.	57
4.4	Carte des sciences naturelles et technologiques d'Ellingham (1948). Avec l'aimable autorisation de la <i>Royal Society</i>	58
5.1	Exemples des deux principaux types de cartes existant en scientométrie.	60

5.2	Représentation cartographique des études zoologiques concernant le taxon Aves en 2008	66
5.3	Représentation cartographique des études zoologiques concernant le taxon Aves en 2008 – Détail : la taxonomie	67
5.4	Représentation cartographique des études zoologiques concernant le taxon Amphibia en 2008	68
5.5	Représentation cartographique des études zoologiques concernant le taxon Amphibia en 2008 – Détail : la taxonomie	69
5.6	Représentation cartographique des études zoologiques concernant le taxon Hemiptera en 2008	70
5.7	Représentation cartographique des études zoologiques concernant le taxon Hemiptera en 2008 – Détail : la taxonomie	71
6.1	Distribution des travaux traitant de Pelobatidae et enregistrés dans le <i>Web of Science</i>	77
6.2	Exemple d'historiographe réalisé avec HistCite	77
6.3	Exemple de référence issue du <i>Web of Science</i> v.4 et exportée au format balisé.	80
6.4	Graphe de citation obtenu à partir de l'historiographe des travaux portant sur les Pelobatidae	81
7.1	Répartition des documents publiés en 2008 et présents dans le <i>Zoological Record</i> suivant les 43 spécialités définies par le premier niveau de son <i>Subject thesaurus</i>	87
8.1	Nombre et proportion de publications pour 5 spécialités du premier niveau du thésaurus du <i>Zoological Record</i>	96
8.2	Proportion des publications zoologiques par spécialité en 1929, 1950, 1980, 1998, 2008	97
8.3	Nombre d'espèces et sous-espèces décrites entre 1864 et 2010 en zoologie	98
8.4	Nombre et proportion d'auteurs de travaux de systématique recensés dans le <i>Zoological Record</i> pour la période 1959-2010	99
8.5	Nombre et proportion de revues publiant des articles de systématique et recensées dans le <i>Zoological Record</i> pour la période 1978-2010 . . .	100
8.6	Nombre d'articles publiant de nouvelles espèces ou sous-espèces, par revue, en 1978 et 2008	101
9.1	Proportion des travaux de systématique par continent et pour les 20 premiers pays de l'auteur de correspondance en 1998 et 2008	105
10.1	Nombre de documents comportant dans leur titre le mot <i>taxonomy</i> dans les journaux <i>Nature</i> et <i>Science</i> entre 1900 et 2010	110

LISTE DES FIGURES

12.1 Fréquence d'utilisation de 26 683 noms d'espèces et sous-espèces distincts dans les documents non systématiciens du *Zoological Record* en 2008 130

Liste des tableaux

2.1	Les cinq types d'indicateurs bibliométriques définis par Vinkler (1988).	13
3.1	Détail des champs AN, C1, BD, DE et ST du <i>ZR</i> .	22
3.2	Champs de recherche disponibles au 10 mars 2010 dans l'onglet de recherche avancée du <i>ZR</i> (Thomson Reuters, 2009)	23
3.3	Les cinq thésaurus du <i>ZR</i> .	24
3.4	Paramètres de recherche temporelle dans le <i>ZR</i> .	28
3.5	Recherches effectuées dans le <i>ZR</i> pour construire le corpus de références sur les Pelobatidae - partie I.	30
3.6	Recherches effectuées dans le <i>ZR</i> pour construire le corpus de références sur les Pelobatidae - partie II	31
3.7	Recherches effectuées dans le <i>ZR</i> pour construire le corpus de références sur les Pelobatidae - partie III.	32
4.1	Proposition de définitions d'un taxonomiste par Haas et Häuser (2006).	50
5.1	Logiciels d'étude de réseaux scientométriques en accès libre	61
5.2	Dénombrement d'auteurs travaillant sur les amphibiens en 2005-2008, par spécialité	65
6.1	Estimation de la couverture taxonomique du <i>Web of Science</i> en 2008	74
6.2	Liste des 10 premières revues publiant des descriptions d'espèces en 2008	75
7.1	Description et période de validité des 7 principaux termes de premier niveau du <i>subject thesaurus</i> du <i>ZR</i> en 2008.	86
8.1	Proportion d'interdisciplinarité en Systématique et dans l'ensemble de quatre autres disciplines en 1929, 1950, 1980, 1998, 2008	92
8.2	Nature des associations multidisciplinaires incluant la Systématique et trois autres disciplines : Evolution, Ecology et Conservation en 1950, 1980, 1998, 2008	92
9.1	Nombre et accroissement du nombre de travaux de systématique entre 1998 et 2008 pour 16 taxons	106

12.1 Liste des 20 espèces les plus fréquemment étudiées dans des études
non taxonomiques en 2008. 129

Pour finir, je ne peux résister au partage de quelques réflexions glanées ici et là au fil de mes lectures, toujours surprenantes et parfois cocasses.

Les travaux de laboratoire et de dissection ne sont pas faciles avec l'extrême chaleur, et quand il faut chercher des plantes ou des animaux dans la campagne, les forces sont bientôt épuisées. Ainsi, la philosophie, le droit, les mathématiques pures conviennent aux pays méridionaux, mais la plupart des sciences physiques et naturelles exigent trop de dépense musculaire pour s'adapter à leurs conditions d'existence. Si l'on peut espérer quelque chose des descendants d'Européens dans les régions tropicales, c'est surtout quand ils vivent à une grande élévation au-dessus de la mer.

(de Candolle, 1873, p. 157-158).

Et pourtant, quoi que j'aie pu faire, je reconnais qu'il est quelques-unes de ces leçons où les faits s'entassent un peu trop, et où la lecture peut devenir fatigante pour l'enfant. Telles les leçons sur la distribution géographique et sur la classification.

(Bert, 1885, p. VI).

Il me semble qu'il est plus honorable d'être le père heureux d'un petit nombre d'espèces bien établies, que le prolifique géniteur de rejetons mal conformés et parus avant terme, destinés le plus souvent à finir leur malfaisante existence dans un vengeur tableau synonymique ou à errer éternellement dans le domaine chaotique et nébuleux des Incerte sedis.

(Racovitza, 1907-1908, p. 148).

One usually finds that taxonomists have wider outlooks and are more generally enlightened than cytologists, excepting, of course, those who are to speak to-day.

(Turrill, 1958).

As the oldest branch of biology, taxonomy appears to be the most set in its ways, but even the foregoing brief citation of some of its responses to changes in biological thought demonstrate that it has by no means reached a stagnant condition. It is nevertheless far less labile than most other branches of biology. This is to be expected considering the vast entourage of genera and species with which it is encumbered on its onward march, not to mention the ghosts of their ancestors clamoring for a place in the parade as they are resurrected from the dead by inquisitive paleontologists.

(Brues, 1939).