

LE STATUT DE LA PARCIMONIE

Guillaume LECOINTRE

UMR CNRS 7138 «Systématique, adaptation, évolution»,
Département «Systématique et évolution», Muséum national d'histoire naturelle,
43 rue Cuvier, 75005 Paris, France.

lecointr@mnhn.fr

Abstract. — Why parsimony in systematics? Parsimony is the operator for measuring the internal consistency (or coherence) of scientific theories. A phylogenetic tree functions as a theory for phylogenetic relationships asserted by hypotheses of transformation. What is to be explained is the distribution of character states among taxa, what explains is phylogenesis, i.e. processes of species diversification, which acts as an explanatory law. The more consistency exists among the hypotheses of character transformation, the less one needs additional hypotheses, and the more parsimonious and powerful the theory (*i. e.*, the tree) is. Thus, the principle of parsimony is nothing but a logical operator of consistency for scientific explanations. It will be shown that this is true using models or not, whatever the epistemology assigned to phylogenetic reconstruction (abduction *versus* hypothetico-deduction). However, the status of parsimony itself often gives rise to interrogations: does it imply that «evolution is parsimonious»? The answer is no: this question arises from a very classical confusion between the properties of our inference tools and the properties of the objects under scrutiny. If parsimony is relevant only to heuristics, why facts predicted by this principle can be confirmed experimentally? It is not surprising that a very consistent phylogeny could be confirmed in an experimental procedure, if this one is consistent itself. This is just due to the fact that parsimony is used in both cases, as it is used in both pattern and process sciences; and has nothing to do with parsimonious natural processes.

Résumé. — Pourquoi la parcimonie en systématique? En sciences, le principe de parcimonie relève de la mesure de la cohérence des théories générées, qu'il s'agisse de la systématique ou de l'histoire. Comment acquérons nous une connaissance historique? Un arbre fonctionne comme une théorie des relations de parenté soutenue par des hypothèses de transformations. Ce qui est à expliquer, c'est la distribution des états de caractère(s) dans la nature. Ce qui explique, c'est la phylogenèse entendue comme processus de diversification des espèces, et qui fait office de loi explicative. La théorie (en l'occurrence l'arbre) est d'autant plus puissante que la proportion de caractères expliqués de manière cohérente, c'est-à-dire sans hypothèses de transformations homoplasiques surnuméraires, est grande. Plus les transformations de caractères sont cohérentes entre elles, et moins il est besoin de faire de telles hypothèses additionnelles. Le C.I. porte bien son nom d'«indice de cohérence». Le principe de parcimonie n'est donc qu'un opérateur de cohérence logique des explications scientifiques. On montre ici que ce statut ne dépend pas du type de parcimonie utilisée en systématique; c'est-à-dire parcimonie «agnostique» *versus* utilisant des modèles explicites d'évolution), ni même de l'épistémologie proclamée; c'est-à-dire hypothético-déduction poppérienne *versus* abduction. Le statut de la parcimonie suscite souvent des interrogations: implique-t-elle que «l'évolution est parcimonieuse»? La réponse est non: cette question émane d'une confusion classique entre les propriétés de nos outils d'inférence et celles des objets analysés, maintes fois dénoncée dans diverses sciences. Si la parcimonie ne relève

que de l'heuristique, pourquoi des faits prédits par le principe de parcimonie rencontrent-ils des confirmations expérimentales ? Obtenir des confirmations expérimentales d'inférences parcimonieuses n'a rien de magique, ni d'étonnant, ni d'informatif sur d'hypothétiques chemins parcimonieux empruntés par la Nature. Il est normal, et même attendu, qu'une phylogénie très cohérente puisse rencontrer une confirmation expérimentale, elle-même cohérente, puisque la parcimonie est utilisée dans les deux cas. Ceci reste possible même si les régimes de preuves de la phylogénie et sciences expérimentales des processus sont différents, car l'usage du principe de parcimonie les transcende tous deux.

SCIENCES EXPÉRIMENTALES, SCIENCES HISTORIQUES

La biosystématique, une science d'historiens ?

Dans tous les secteurs de la connaissance objective, il existe une coupure épistémologique profonde entre les sciences des structures et les sciences des processus. Elles ne se distinguent pas par leurs sources de données – qui dans les deux cas peuvent être mixtes, relevant à la fois des structures et des processus –, ni par les objets qu'elles analysent, mais par le type de question ultime. Pour schématiser à l'extrême, les premières répondent à la question « quoi ? » et les secondes à la question « comment ? » (table 1). On appelle induction une généralisation empirique directe, c'est-à-dire le passage d'instances observées à une nouvelle instance non observée. On appelle aussi induction le passage d'un nombre limité de cas observés à une relation causale (ou loi causale) généralisée. Même si l'induction est largement pratiquée dans toutes les sciences, on peut dire que dans les deux secteurs, sciences des structures et sciences des processus, la façon dont on prouve tend à être différente. En effet, on peut trouver des sciences des structures qui pratiquent l'induction, (la taxonomie, par exemple), et des sciences des processus qui font de même (la psychologie, par exemple). On peut trouver des sciences des structures pratiquant l'hypothético-déduction stricte, prédictive, par exemple la cristallographie, qui dispose de lois qui, à partir de prémisses, impliquent *nécessairement* le résultat. On peut trouver des sciences des processus qui font de

même, la physiologie, par exemple, qui dispose parfois de lois formalisées. Cependant, les sciences des structures pratiquent majoritairement l'induction, simplement parce qu'à l'observation d'un nombre limité d'instances s'ensuit une généralisation d'un attribut à des entités plus grandes. L'hypothético-déduction semble davantage – tout au moins en Biologie – l'apanage des sciences des processus, dont les phénomènes sont, au mieux, décrits en tout ou partie pas des lois prédictives. Qu'en est-il de la systématique phylogénétique ? Pour MAHNER & BUNGE (1997) toute classification est conventionnelle. Sa pertinence se borne à respecter au mieux un cahier des charges. Depuis la Renaissance, les sciences des structures du Vivant produisent des classifications dont la vocation n'est plus d'être utilitaire, mais d'être « naturelle ». Depuis Darwin, le cahier des charges d'une classification naturelle est d'être « généalogique », nous disons aujourd'hui « phylogénétique ». L'arbre phylogénétique procède à une mise en ordre relative des acquisitions d'attributs dans le temps. En ce sens, il ne raconte en toute première instance ni vraiment une histoire, ni un scénario, mais un *déroulement* toujours partiel. C'est ensuite qu'il deviendra une classification phylogénétique, selon le cahier des charges qui est de rigueur pour toute classification moderne du vivant. Le régime de la preuve n'est ni vraiment complètement de l'induction (bien que des généralisations soient produites *a priori* dès la matrice et postérieurement à l'arbre, *a posteriori*), ni de l'hypothético-déduction au sens strict : il n'y a pas de loi qui, à partir de prémisses, implique *absolument* un résultat, de telle sorte qu'elle permettrait de faire des prédictions. Certains ont proposé qu'en systématique, le fameux « descent with modification » fasse office de loi (FARRIS, 1983). Cette proposition est difficilement tenable si le mot « loi » est pris au sens des lois déductives prédictives de la physique ou de la chimie. La filiation avec modification, qui est une façon de parler du processus de phylogenèse, est un aspect de la théorie générale dans le cadre de laquelle nous travaillons, une explication générale de toute répartition des caractères sur tout arbre. La phylogenèse fait bien office de loi explicative, mais une

loi « molle » qui, au lieu de rendre un résultat certain à partir de prémisses, ne rend ce résultat que probable. Le régime de preuve de ce qu'il est convenu d'appeler la « biosystématique » s'apparente en fait à celui utilisé par les historiens (table 1 : l'abduction est le régime de preuve de la biosystématique). Les degrés d'apparement d'une phylogénie ne sont pas obtenus expérimentalement, au sens où on récolterait directement les données à l'aide d'une machine à remonter le temps, ni sur la base de registres d'état civil, ni à l'aide d'une loi formalisée prédictive. Ces arbres résultent d'un exercice de reconstitution d'un passé à partir d'observations d'éléments présents à expliquer. Nous devons à FITZHUGH (1997, 1998) et GEIGER *et al.* (2001) le rapprochement de cet exercice avec un raisonnement abductif. L'abduction est un type d'inférence utilisant une loi qui, à partir de prémisses (l'état final d'un système), rend le résultat (ses états antérieurs) seulement *probables*. L'abduction est une déduction non pas prédictive, mais rétrodictive, une déduction à rebours, en quelque sorte. La phylogénétique (« biosystématique ») est abductive parce qu'elle infère un déroulement partiel d'évènements passés qui ont donné naissance à la hiérarchie des attributs que nous observons aujourd'hui. Qu'est-ce que cette hiérarchie ? Prenons le cas de la parcimonie « minimale », « standard », ignorante ou agnostique vis-à-vis de tout modèle d'évolution ou de toute pondération. Si nous avons cinquante espèces animales devant les yeux, nous sommes immédiatement capables d'observer leurs attributs. Certaines ont quatre pattes. Parmi celles-ci, certaines ont des poils. Parmi celles-ci, certaines ont le pouce opposable au reste des doigts. Ces attributs (pattes, poils, pouce opposable) ne sont pas distribués n'importe comment : tout ceux qui ont le pouce opposable ont déjà les poils, tous ceux qui ont des poils ont déjà quatre pattes... C'est-à-dire que la répartition des attributs n'est pas totalement désordonnée : on ne trouve pas de poils en dehors de ceux qui ont quatre pattes, ni de pouce opposable en dehors de ceux qui ont des poils. Il y a des attributs à expliquer, leur mise en cohérence maximale se traduit par la construction d'un arbre. Ici, la cohérence maximale

Table 1. Les sciences vues au travers des types de questions posées et de leur régime de preuve. Dans les cases sont donnés des exemples de sciences. On voit que la systématique (taxonomie et biosystématique) est avant tout une science des structures du vivant. La taxonomie est inductive parce qu'elle généralise au taxon les attributs observés sur quelques exemplaires. La phylogénétique (« biosystématique ») est abductive parce qu'elle infère un déroulement partiel des évènements qui ont donné naissance à la hiérarchie des attributs que nous observons aujourd'hui. Comme exemple de science des structures utilisant un régime de déduction, nous choisissons la cristallographie, car ce sont des lois qui régissent la déduction des structures tridimensionnelles des cristaux.

	Structures	Processus
Induction	Taxonomie	Psychologie
Abduction	Biosystématique	Histoire
Déduction	Cristallographie	Physiologie, Biochimie...

consiste à mettre dans un seul et même clade tous ceux qui ont des poils, au lieu de les ranger séparément avec ceux qui n'en ont pas dans des clades distincts. De manière sous-jacente à notre action (mais non inscrite *in silico* dans cette version minimale de l'utilisation de la parcimonie), c'est la phylogénèse qui explique cet emboîtement des attributs. Signalons au passage que cet « ordre naturel », cette hiérarchie, le plus souvent n'est pas perceptible d'emblée aussi facilement, et c'est la raison pour laquelle nous faisons appel à un algorithme de mise en cohérence. L'arbre phylogénétique résultant traduit *à la fois* les degrés relatifs d'apparement des espèces par l'emboîtement de leurs attributs, et le déroulement historique le plus cohérent de ces attributs, c'est-à-dire l'ordre relatif de leur acquisition dans le temps. On a donc reconstitué un *déroulement argumenté et vérifiable par autrui*. Cet ordre d'apparition n'est que le déroulement le plus crédible parmi d'autres. Certains diraient même le plus « probable », comme RIEPPEL (2003 ; Pers. Com.), qui parle de « *probabilification of systematics* ». Cet ordre n'est pas absolument requis par une loi formalisée. La loi explicative, la phylogénèse, est dépourvue de tout formalisme prédictif. L'historien ne procède pas différemment, mais il ne dispose pas d'algorithme pour produire la mise en cohérences

des données. Il n'y a pas non plus de machine à remonter le temps pour observer la bataille d'Austerlitz, ni de témoins oculaires vivants. On pense qu'elle a existé parce que des faisceaux de présomption, des indices et autres données (documents, objets présumés relever de cet événement) sont d'une grande cohérence, et qu'il n'existe pas de théorie alternative plus cohérente encore. La bataille d'Austerlitz est une théorie hautement corroborée dont on peut décrire un *déroulement*, et dont la mise en cohérence des observations est ni formalisée ni informatisée, ce qui fut longtemps le cas également pour les théories des biosystématiciens.

La preuve expérimentale

La preuve expérimentale, quant à elle, consiste à agir ici et maintenant sur le monde réel en produisant une expérience décisive qui va permettre de trancher entre plusieurs hypothèses en concurrence. En biologie de l'évolution par exemple, on va mimer des forces évolutives telles qu'on se les représente. Pour simuler l'origine abiotique de molécules biologiques tels les acides aminés, STANLEY MILLER et HAROLD UREY ont soumis des composés abiotiques simples (méthane, hydrogène, ammoniaque, eau) à certaines conditions physiques dont on pensait qu'elles devaient être celles d'une terre primitive (chaleur, électricité). Ils ont fabriqué *in vitro* de nombreux acides aminés (constituants élémentaires des protéines) et les bases puriques des acides nucléiques (constituants élémentaires de l'ADN). Lorsqu'ils travaillaient sur des espèces à temps de génération très courts, les biologistes ont pu « voir » l'évolution dans leur laboratoire. Dès les années trente, PHILIPPE L'HÉRITIER et GEORGES TEISSIER ont vérifié l'évolution biologique expérimentalement en maintenant des populations de 3 000 à 4 000 petites mouches du vinaigre dans des cages et en les soumettant à des contraintes variées, notamment alimentaires. On fait aujourd'hui cela couramment avec des bactéries ou des phages, notamment lors de « phylogénies expérimentales » réalisées en laboratoire (HILLIS *et al.*, 1992). Ici c'est l'expérience qui explique la phylogénèse (même s'il s'agit surtout de *mimer* en laboratoire ses mécanismes possibles

et de valider des modèles). Soulignons au passage que toutes les disciplines de la biologie ne miment pas nécessairement l'évolution. L'hypothético-déduction est illustrée au mieux lorsque la preuve expérimentale de processus à l'œuvre dans le vivant utilise – bien sûr – des lois déductives au pouvoir prédictif. De telles lois sont trouvées en physiologie, en biochimie, en biomécanique, par exemple. Ici point de rétrodictio. L'expérience menée vise à tester une hypothèse à l'aide de lois qui, à partir de prémisses, impliquent nécessairement le résultat et permettent des prédictions.

Se comprendre

Il faut éviter de croire que toutes les sciences prouvent de la même façon. Juger d'une autre science en projetant son propre régime de preuve sans avoir vérifié ce qui y était à l'œuvre est une erreur hélas courante. Les sciences des structures (anatomie comparée, embryologie descriptive, paléontologie, biosystématique incluant la phylogénie moléculaire...) sont des sciences historiques : la phylogénèse – c'est-à-dire le processus à l'origine de la « filiation avec modification » de Darwin – *explique* la répartition des structures à travers le vivant. Les sciences des processus (génétique moléculaire, embryologie causale, physiologie, écologie...) sont des sciences expérimentales où la phylogénèse *est expliquée* par des expériences. Dans le premier cas, la phylogénèse explique, dans le second elle est à expliquer. Si l'on se trompe de régime de preuve, on arrive vite à des raisonnements circulaires, des condamnations ou des complexes. Ni complètement inductive, ni vraiment hypothético-déductive (au sens des prédictions poppériennes), la systématique est souvent mal comprise. Certains scientifiques prétendent que la systématique n'est pas une science parce qu'ils ont l'intuition qu'elle ne suit pas un schéma argumentatif de type hypothético-déductif fondé sur une expérience sur le monde réel disposant de lois prédictives. C'est sans doute un effet secondaire de l'explosion de la biologie au XX^e siècle, qui est surtout l'explosion des sciences des processus. En contrecoup, certains systématiciens ont absolument cherché à faire entrer de force la biosysté-

matique dans un moule hypothético-déductif, et continuent à le faire (KLUGE & WOLF, 1993 ; GRANT & KLUGE, 2003).

ERREUR CLASSIQUE SUR LA PARCIMONIE

L'arbre est une théorie des relations de parentés qui explique la distribution des états de caractères dans la Nature par des hypothèses de transformations. Le passage d'un état de caractère à un autre sur une branche ne prend son sens de transformation évolutive qu'à l'aide du recours sous-jacent à une théorie plus générale encore, celle de l'évolution, et plus particulièrement une loi explicative, la phylogénèse, ou filiation des organismes avec modification et transmission de leurs attributs. L'arbre parcimonieux est donc l'explication la plus cohérente des états de caractères sous cette loi (voir DELEPORTE, 2004, ce volume). Ceci reste vrai quelle que soit l'école affichée : que la théorie générale ne s'inscrive que dans l'intention du chercheur ou dans le métalangage utilisé *a posteriori* pour parler du degré de corroboration d'un résultat phylogénétique, ou bien que la théorie générale et la loi s'inscrivent *in silico* par l'utilisation de modèles d'évolution ou de pondérations sophistiquées. Répétons-le, ce qui est à expliquer, ce sont les similitudes entre attributs. Ce qui explique, c'est la phylogénèse, car elle fait office de loi explicative. Pour un jeu donné de caractères, il existe un nombre très grand d'arbres possibles, de cohérences variables quant à la contiguïté des attributs semblables qu'ils impliquent. Pourquoi choisissons-nous le plus parcimonieux ? Un arbre est une micro-théorie impliquant des hypothèses de transformations de caractères. Les théories (et donc les arbres) en concurrence se mesurent à l'aune de leur cohérence interne. Par chance, en reconstruction phylogénétique, nous avons les moyens de mesurer la cohérence de nos théories, ce que n'ont pas les historiens qui utilisent le même régime de preuve mais sans algorithmes. La parcimonie est juste une mesure de la cohérence des arbres, c'est-à-dire qu'elle mesure la cohérence des données sous une loi explicative. D'une manière générale, une théo-

rie est cohérente lorsque les hypothèses qu'elle implique sont compatibles ou se corroborent entre elles. Plus elles se corroborent et moins il est besoin d'en faire. L'arbre le plus parcimonieux est simplement le plus cohérent, maximisant la congruence des caractères et minimisant la fraction contradictoire des caractères qui sont alors interprétés sous forme d'hypothèses *ad hoc* d'homoplasies (Farris, 1983), toujours sous une loi explicative particulière.

Par conséquent, l'utilisation de la parcimonie n'est pas liée à l'idée d'un processus évolutif qui serait parcimonieux. Cette erreur courante provient d'une confusion classique dans toutes les sciences, et qui consiste à confondre les propriétés des concepts et des outils d'investigation avec les propriétés des objets que l'on étudie. Certains auteurs relativisent l'utilisation de la parcimonie en clamant que l'évolution n'est pas nécessairement parcimonieuse. Citons, à titre d'exemple, ANDERSON (1994, p. 2) : « *Evolution, however, does not always proceed parsimoniously, and to base phylogenies solely on this criterion is simply dogmatic* ». D'autres semblent confondre rareté (notamment celle des mutations, TASSY, 2000, p. 41, bien que les mutations ne soient pas rares) et parcimonie. TASSY (2000, p. 41) s'interroge sur l'analogie entre un minimum mathématique interprété comme « principe d'économie » suivi par la Nature et parcimonie. Le minimum mathématique n'est minimal que par la maximisation mathématisée de la cohérence des faits dont on rend compte. La bulle de savon, en enfermant le maximum de volume dans un minimum de surface, ne saurait être parcimonieuse. L'analogie ne tient pas : ce sont nos outils mathématiques qui, tout en rendant compte de ce volume comme maximal et de cette surface comme minimale, appliquent un critère de maximisation de la cohérence de ce qui est observé, autrement dit un critère de parcimonie. D'autres encore s'interrogent sur la crédibilité présumée (mesurée comment ?) de ce qu'implique la parcimonie en nombre d'événements évolutifs sur un problème biologique donné. Ces derniers auraient meilleur avantage à intégrer les connaissances qui les font douter soit directement dans la matrice, soit dans un modèle d'évolution ou un schéma de pondé-

ration. Ces dévoiements du principe de parcimonie ressemblent à l'erreur du Cardinal Paul Poupard qui justifie l'existence de Dieu parce que la Nature serait « rationnelle » (POUPARD, 1994). Tout comme la rationalité n'est pas une propriété de la Nature (ce qui n'a aucun sens) mais bien une faculté de l'observateur, et qui lui sert d'outil d'inférence, la parcimonie n'est pas une propriété ni de la Nature ni de l'évolution mais un moyen de rendre les plus cohérentes possibles nos inférences. Comme l'écrit TASSY (2000), mais malheureusement sur le mode interrogatif : « *Le principe de parcimonie n'est-il qu'un principe méthodologique pour forcer le réel à exister en même temps qu'il devient intelligible ?* »

QUE PENSER DES CONFIRMATIONS EXPÉRIMENTALES ?

Pourquoi la parcimonie, si elle n'est qu'un principe d'investigation, voit-elle ses inférences confirmées expérimentalement ? Plusieurs études ont montré que la parcimonie retrouvait les résultats attendus dans le cadre d'évolutions simulées *in silico* (HUELSENBECK, 1995) ou bien *in vitro* sur des organismes à temps de génération court (HILLIS *et al.*, 1992). Cela ne montre-t-il pas qu'il y a un rapport entre l'outil d'inférence et la réalité du processus à l'œuvre ? La réponse est non. Obtenir des confirmations expérimentales de faits attendus par une approche parcimonieuse n'a rien de magique, ni d'étonnant, ni d'informatif sur d'hypothétiques chemins parcimonieux empruntés par la Nature. Cela provient du fait que l'utilisation du principe de parcimonie, c'est-à-dire, répétons-le, de maximisation de la cohérence des données selon la loi explicative, formalisée ou non, est beaucoup plus générale en science que ce que les systématiciens imaginent. Il est normal et même attendu qu'une phylogénie très cohérente puisse rencontrer une confirmation expérimentale, elle même rendue cohérente. Ceci reste possible même si les régimes de preuves entre phylogénie d'une part, et sciences expérimentales des processus d'autre part, sont différents, car l'usage du principe de parcimonie les transcende tous les deux.

MISES EN COHÉRENCES PLUS COMPLEXES

La version de la parcimonie qui est exposée ci-dessus dans la section 1 correspond à son utilisation standard, en absence d'éléments de connaissances externes à ce qui est déjà codé dans la matrice. Cette absence d'éléments supplémentaires correspond à deux situations épistémologiques contrastées. Soit des connaissances externes sont absentes lorsqu'on a mis le maximum des connaissances disponibles déjà dans la matrice. Ceci peut se produire par exemple lorsque l'on a codé dans une matrice tout ce qu'on sait des caractères morphologiques disponibles d'un groupe d'organismes. Soit ces connaissances externes sont disponibles mais les auteurs les ignorent délibérément, pensant que moins on les prend en compte, plus on renforce la puissance du « test ». Cette notion de test va de pair avec un hypothético-déductivisme poppérien proclamé. Par exemple, les adeptes du « coherentism » (selon RIEPPEL, comm. pers.), dans leur pratique du « total evidence » (KLUGE, 1989), refusent d'apporter un quelconque crédit au fait que plusieurs gènes indépendants sont inclus dans la matrice (SIDALL, 1997), et que la délimitation de chacun n'est pas qu'un artefact technique ou historique, mais correspond à une unité fonctionnelle intégrée (contrairement à un jeu de caractères morphologiques hétérogènes). J'ai été moi-même de leur avis (LECOINTRE, 1997) avant d'avoir exploré plus en profondeur à quelle épistémologie se référerait KLUGE (1989) (mais voir LECOINTRE & DELEPORTE, 2004). Ils refusent également, par exemple, la connaissance du code génétique qui nous enseigne que l'espace mutationnel est très restreint pour les transitions de première position du codon. Ces dernières connaissances n'ont pourtant rien de phylogénétique, et n'entraîneraient aucunement un défaut d'indépendance de l'analyse (DELEPORTE, 1993), mais seulement une dépendance de celle-ci vis-à-vis de ces connaissances de base. N'en déplaise aux partisans de l'indépendance de la systématique ou des « *null or minimal models* » (BRADY, 1985 ; PATTERSON, 1994), il est d'une grande naïveté de penser qu'il puisse y avoir des « faits » ou même des objets intelligibles sans théo-

rie pleinement assumée qui leur donne sens, et qui, en phylogénie, passe par un modèle d'évolution, implicite ou rendu explicite. Si ces connaissances sont ignorées, on perd du même coup la possibilité de détecter des artefacts, par exemple par croisement de la connaissance relative à l'espace mutationnel et l'analyse de la congruence taxonomique (sans arbres consensus). Quoi qu'il en soit, même dans une parcimonie proclamée « agnostique », c'est-à-dire sans modèle, celle-ci correspond à la maximisation de la cohérence des données (sous un modèle *non reconnu*, celui d'une absence d'évolution concertée entre caractères). KLUGE, grand défenseur d'une parcimonie agnostique, fait de la parcimonie une opération d'optimisation (c'est-à-dire de mise en cohérence) sous un critère donné (qui n'est qu'une expression de la loi explicative) : « *As I see it, parsimony is simply the operation of optimizing on a criterion* » (KLUGE, 1994, p. 354). Dommage qu'il ne nous éclaire pas davantage sur la façon dont le « background knowledge » qu'il refuse nourrit ce critère. Mais ceci est un autre débat.

Mais la pratique de la parcimonie peut aussi prendre en compte toutes les connaissances disponibles dans des jeux de pondération, dans l'acceptation de partitions biologiquement justifiées et dans l'analyse de la congruence taxonomique entre arbres issus de ces parties des données prises séparément. Les connaissances prises en compte réalisent une mise en cohérence à double détente : une première au niveau des caractères de chaque matrice, puis une seconde entre les résultats issus de chacune de celles-ci. Nous avons montré par ailleurs que le gain résultant de la prise en compte des connaissances disponibles est une meilleure évaluation de la fiabilité des inférences phylogénétiques (LECOINTRE & DELEPORTE, 2004), si la notion de fiabilité n'est pas collée à celle de robustesse statistique, mais considérée comme un métalangage qui porte sur le degré de confiance que l'on accorde à un propos (la fonction de confirmation de CARNAP (1950)). Dans un tel cadre, la parcimonie se pense obligatoirement avec un modèle, qui est la traduction d'une partie de la loi explicative. Il n'y a pas de test poppérien ni de réfutation : on maximise la cohérence des propos

sous une loi explicative. Lorsqu'il n'y a pas de modèle d'évolution commun à un sous ensemble de caractères, ou pas de connaissances externes permettant d'identifier des partitions indépendantes et biologiquement justifiées, la parcimonie opère alors sous un modèle *reconnu* qu'on pourrait nommer « un modèle par caractère » ou « pas d'évolution concertée des caractères ». Mais cela ne change rien à son statut. Elle reste une fonction de maximisation de la cohérence des données selon la loi explicative (armée d'un modèle explicite). Pour traduire ces termes épistémologiques dans la pratique du systématicien, nous dirons que la parcimonie est une fonction de maximisation de la cohérence des caractères selon un modèle d'évolution, implicite ou explicite, pauvre ou riche, tiré de la phylogénèse jouant le rôle d'une loi. Dans ce cadre, elle n'est pas, non plus, une propriété de la Nature.

La parcimonie, comprise comme principe général de maximisation de la cohérence, pourrait donc être utilisée selon diverses modalités. À quel niveau réalise-t-on la maximisation de la cohérence ? L'attitude relevant du « coherentism » ne reconnaît aucun modèle et n'admet que la cohérence de tous les caractères mis ensemble. Elle fonde la fiabilité sur la robustesse calculée sur une seule matrice. C'est l'école du « total evidence » (KLUGE, 1989 ; KLUGE & WOLF, 1993 ; GRANT & KLUGE, 2003). L'attitude relevant du « foundationalism » (RIEPEL, comm. pers.), ayant abandonné la négation hypothético-déductiviste de tout « background knowledge », accepte l'idée contraire que le degré de confirmation d'une théorie incorpore sa cohérence avec le « background knowledge » pleinement assumé. Nous (LECOINTRE & DELEPORTE, 2000 ; 2004) avons même établi en quoi cette pratique était conforme à la notion de « total evidence » telle que la concevait CARNAP (1950). La cohérence est alors assurée en incorporant des modèles explicites d'évolution dans la recherche de l'arbre, ce qui se fait dans l'analyse de séquences nucléotidiques soit par maximum de vraisemblance, soit en parcimonie pondérée. À un niveau supérieur, la cohérence est assurée par l'analyse de la congruence taxonomique, et dans ce cas la fiabilité peut se mesurer au nombre de fois où des

données indépendantes conduisent au même clade (c'est, en fait, le métalangage scientifique le plus courant). Ainsi les approches probabilistes (par exemple de maximum de vraisemblance) et les méthodes de parcimonie pourraient être mises sous la même bannière épistémologique. Dans tous les cas, la parcimonie reste une fonction de maximisation de la cohérence des caractères selon un modèle d'évolution, implicite ou explicite, sous une loi explicative, la phylogenèse. La parcimonie est l'opérateur de la cohérence logique de toute explication scientifique.

Remerciements. — Ayant écrit cet article sur un bateau très loin de mon bureau, je remercie Pierre Deleporte pour le dépannage bibliographique, et pour les précieux commentaires qu'il fit à cette occasion. — Je remercie la compagnie Saint-Gobain pour l'aide apportée à mes travaux d'épistémologie.

BIBLIOGRAPHIE

- ANDERSON E.M., 1994. Systematics and osteology of the Zoarcidae (Teleostei : Perciformes). *Ichthyological Bulletin* n° 60, J.L.B. Smith Institute of Ichthyology, Grahamstown, South Africa.
- BRADY R., 1985. On the independence of systematics. *Cladistics* 1 : 113-126.
- CARNAP R., 1950. *Logical foundations of probability*. Univ. Chicago Press, Chicago.
- DELEPORTE P., 1993. Characters, attributes and tests of evolutionary scenarios. *Cladistics*, 9 : 427-432.
- DELEPORTE P., 2004. Parcimonie ou maximum de vraisemblance : mieux considérer les postulats pour en finir avec une querelle de sourds. *Biosystema*, 22 (ce volume).
- FARRIS S., 1983. The logical basis of phylogenetic analysis. In : N. Platnick & V. Funk (eds.), *Advances in cladistics* 2, p. 7-36, Columbia Univ. Press, New York.
- FITZHUGH K., 1997. The abduction of cladistics. *Cladistics*, 13 : 170-171.
- FITZHUGH K., 1998. C (h, be) : e = / = Synapomorphy. In : *Abstracts XVIIIth meeting of the Willi Hennig Society, Sao Paolo*, p. 34-35.
- GEIGER D., FITZHUGH K. & THACKER C., 2001. Matters of the record. Timeless characters : a response to Vermeij (1999). *Paleobiology*, 27 : 179-180.
- GRANT T. & KLUGE A., 2003. Data exploration in phylogenetic inference : scientific, heuristic, or neither. *Cladistics*, 19 : 379-418.
- HILLIS D.M., BULL J.J., WHITE M.E., BRADGETT M.R. & MOLINEUX I.J., 1992. Experimental phylogenetics : generation of a known phylogeny. *Science*, 255 : 589-592.
- HUELSENBECK J.P., 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.*, 44 : 17-48.
- KLUGE A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38 : 7-25.
- KLUGE A., 1994. Sophisticated falsification and research cycles : consequences for differential character weighting in phylogenetic systematics. *Zoologica Scripta*, 26 : 349-360.
- KLUGE A.G. & WOLF A.J., 1993. Cladistics : what's in a word? *Cladistics*, 9 : 183-199.
- LECOINTRE G., 1997. Dialogue généticiens/systématiciens : une culture du quantitatif contre celle du qualitatif? *Biosystema*, 15 : 7-20.
- LECOINTRE G. & DELEPORTE P., 2000. Le principe du « total evidence » requiert l'exclusion de données trompeuses. *Biosystema*, 18 : 129-151.
- LECOINTRE G. & DELEPORTE P., 2004. Total evidence requires exclusion of phylogenetically misleading data. *Zool. Scripta*, 33 : 1-17.
- MAHNER M. & BUNGE M., 1997. *Foundations of biophilosophy*. Springer-Verlag, Berlin.
- PATTERSON C., 1994. Null or minimal models. In : R.W. SCOTLAND, D.J. SIEBERT & D.M. WILLIAMS (eds.), *Models in Phylogeny reconstruction*, Systematics Association Special Volume, 52, p. 173-192, Clarendon Press, Oxford.
- POUPARD P., 1994. *Après Galilée. Science et foi : nouveau dialogue*. Desclée de Brouwer, Paris.
- RIEPEL O., 2003. Popper and systematics. *Syst. Biol.*, 52 : 259-270.
- SIDDALL M.E., 1997. Prior agreement : arbitration or arbitrary? *Syst. Biol.* 46 : 765-769.
- TASSY P., 2000. La théorisation du caractère morphologique en cladistique. *Biosystema*, 18 : 33-43.