Pour une ontologie formelle et pluriformalisée de la géographie.

P. Langlois, UMR CNRS IDEES – université de Rouen

Le premier objectif décrit dans ce texte, peut être la base d'une proposition de programme de recherche dans « l'axe » géographie de COS-MA-GEM.

Ce texte est pour l'instant personnel, mais je propose que chacun des géographes (et aussi les autres) se positionnent par rapport à la lui, pour faire évoluer la proposition, ou me dire si c'est une mauvaise piste…

Notre premier objectif est de construire une théorie générale des objets géographiques. Ce travail constitue une ontologie formelle de la géographie, dans le sens où notre objectif n'est pas de décrire la liste des objets géographiques existants, (lieu, territoire, frontière, paysage, bassin versant, bâtiment, réseau routier, ville, aire de chalandise, front pionnier, etc.) mais d'en dégager les classes, les structures, les principes et propriétés générales, la manière dont les objets géographiques s'organisent, la manière dont ils fonctionnement, les opérateurs qui permettent de les faire évoluer, de les transformer, de les combiner etc. Pour qu'il n'y ait pas de confusion entre la catégorie sémantique des noms et celle des objets, nous formalisons de plus cette ontologie à la fois mathématiquement et informatiquement en l'associant à l'espace de noms de la géographie qui permet de lui conférer une sémantique. Après avoir explicité « ce qui est », comment se structure et s'organise l'univers de la réalité géographique, notre second objectif est de concevoir et d'implémenter cette vision théorique dans un système technologique forcément réducteur, mais capable de réifier cette ontologie formelle à travers une classe de modèles exécutables adaptée aux problématiques des systèmes complexes en géographie.

Pour réaliser cette ontologie formelle, nous pensons que cette démarche théorique doit être formalisée mathématiquement car cela permet d'élaborer une construction totalement indépendante de toute contrainte technologique et fournit un cadre théorique rigoureux éprouvé et reconnu. Cela permet donc, dans un premier temps, de penser les objets géographiques selon une démarche tournée vers la compréhension de la réalité, même si elle reste simplificatrice. Les outils mathématiques permettent par exemple de penser le continu ou l'infini, sans se préoccuper des limitations de l'ordinateur, dans lequel finalement, tout doit être explicité, tout est fini, tout est discret. Nous avons besoin d'un cadre théorique pour formaliser cette construction. D'abord, la théorie des ensembles et de la logique des prédicats forment un socle élémentaire aujourd'hui reconnu pour toute formalisation mathématique. Même si les travaux de Gödel sur l'incomplétude, ruinèrent le projet de Hilbert de constituer un fondement totalement formalisé et cohérent des mathématiques et réfuta les thèses formalistes du Cercle de Vienne, nous sommes loin des grandes interrogations des débuts de cette théorie initiée par Cantor au début du XIX^{e} siècle, (Russel, Frege), où des paradoxes fondamentaux ébranlèrent son édifice axiomatique. La théorie des ensembles a conquis sa maturité en prenant conscience de ses limites, par exemple en sachant faire la distinction entre ce qu'est un ensemble et ce qui n'en est pas un (qu'on appellera alors famille ou collection). La notion d'ensemble ne se définit pas formellement, c'est une définition première de la théorie. Néanmoins la famille de tous les ensembles n'est pas un ensemble, car un ensemble doit être défini clairement, soit par la liste exhaustive et non redondante de ses éléments, (il est alors défini en extension), soit par une propriété logique caractéristique de ses éléments, (il est alors défini en compréhension). Une autre règle essentielle pour ne pas que la théorie soit contradictoire porte sur la relation d'appartenance : un ensemble ne peut pas appartenir à lui-même. Par contre, la notion de sous-ensemble (qu'on appelle aussi une partie d'un ensemble), donne naissance à la relation d'inclusion qui est une relation d'ordre définie sur l'ensemble $\mathcal{P}(E)$ des parties d'un ensemble E. La relation d'inclusion est réflexive, contrairement à celle d'appartenance qui est antiréflexive. Ainsi, en théorie des ensembles, un ensemble se contient toujours lui-même mais n'appartient jamais à lui-même. Avec ces précautions, le paradoxe de Russel n'existe plus. En effet, ce paradoxe reposait sur un « ensemble » particulier, formé de tous les ensembles qui ne se contiennent pas eux-mêmes. Le paradoxe venait du fait que cet « ensemble » ne pouvait, ni se contenir lui-même ni, ne pas se contenir. Ces améliorations étant réunies, avec d'autres, dans l'axiomatique de Zermelo-Fraenkel, elle donne à cette théorie une grande solidité. Même si elle n'est pas la seule qui soit à la base d'une théorie des ensembles, c'est celle qui est très largement utilisée aujourd'hui, accompagnée éventuellement d'autres axiomes complémentaires comme l'axiome du choix ou l'hypothèse du continu. Quelques autres théories des ensembles ont été formalisées, comme la théorie des types (Whitehead, Russell) ou la théorie des classes (von Neumann, Bernays, Gödel). Malgré leurs différences, ces théories apparaissent maintenant comme des traductions convergentes d'une même réalité mathématique. D'autres tentatives d'axiomatisation se sont développées dans des directions plus éloignées dont certaines ont été formalisées. C'est le cas de la méréologie qui est plutôt une théorie logique, formalisée par le logicien Stanisław Leśniewski (1886–1939). Il nous semble qu'elle ne constitue pas une avancée plus féconde pour notre travail que la théorie des ensembles « standard ». Par exemple, un des principes fort du paradigme de la complexité, est que le tout est plus que ses parties. En théorie des ensembles, comme en méréologie, cette affirmation est fausse. La définition d'un système complexe doit donc reposer sur un concept plus riche qu'un simple ensemble formé d'éléments (et de parties). C'est cet enrichissement que nous proposons de construire, qui n'est en rien contradictoire avec l'utilisation de la théorie des ensembles. De plus, cette construction ontologique ne se borne pas à la seule utilisation de la théorie des ensembles, tout l'édifice de l'algèbre, de la géométrie, de la topologie et de l'analyse, dont la cohérence et le langage s'appuient sur la théorie des ensembles, pourra nous être utile à divers degrés. Néanmoins, nous ne cherchons pas ici à élaborer une nouvelle théorie mathématique, formée d'une suite de théorèmes et de démonstrations, ni de nouveaux axiomes, mais à utiliser son langage, ses objets, sa cohérence, sa non-ambigüité, pour énoncer les définitions des entités de notre ontologie. Ainsi, « le niveau » de connaissances mathématique utilisé restera élémentaire.

Pour préciser cette démarche de construction ontologique du domaine géographique, on part du « néant » des localisations, formalisé par « l'espace »[1] de la géométrie, vide de toute matière, de tout contenu. Ce néant permet de construire des formes géométriques et de les mettre en relation à travers la topologie pour construire des objets abstraits plus complexes. Mais l'essence profonde des objets n'apparaît qu'avec l'adjonction à l'espace du concept de matière-énergie. Comment la matière prend-elle sa réalité dans le néant géométrique ? Un point, une ligne ou une surface existent-ils encore lorsque l'espace devient matériel? Enfin, la prise en compte du temps permet de construire des faits et des comportements, de faire naître, se développer et mourir des êtres (physiques, vivants, sociaux, imaginaires) et ajouter à la

[1] Nous employons ce terme de manière volontairement ambiguë pour évoquer à la fois le domaine disciplinaire qu'est la géométrie, mais aussi pour indiquer qu'on se situe dans un espace physique formalisé mathématiquement.

diversité de la réalité spatiale, la profondeur de l'histoire et l'incertitude des devenirs. Le triptyque espace-temps-matière est le cadre conceptuel général de cette construction ontologique. De plus, les concepts d'agent et d'organisation sont au cœur de la construction des géographiques. Ils définissent un être dual. Constitué d'une part de la face externe de son enveloppe, qui lui permet d'agir vers le monde qui l'environne et qu'il perçoit comme formé d'une diversité d'autres objets-agents de la même conception. Constitué d'autre part de la face interne de son enveloppe, qui regarde l'objet comme une organisation tournée vers la profondeur récursive de son organisation intérieure, dont les composants sont encore des objets-agents formant système. Ces deux facettes expriment l'interaction fondamentale, l'essence même de l'objet, où s'affrontent les comportements venant des profondeurs de l'être contre ceux nécessités par les réalités extérieures. Cette ontologie sera appelée (Agent-Organisation-Comportement) ou AOC. Elle semble assez proche de la démarche AGR (Agent-Groupe-Rôle) de Jacques Ferber, bien que cette dernière ne contienne pas cette dualité récursive de l'objet comme agent vu de l'extérieur et organisation vu de l'intérieur et n'intègre pas les environnements externe et interne qui en découle. La raison essentielle de cette structure vient du fait qu'elle exprime bien une vision systémique multi-échelle, importante en géographie. De plus, elle permet de bien identifier les limites à la fois extérieures et intérieures d'un modèle. Ainsi, on peut souvent identifier trois niveaux de modélisation (mais ce nombre n'est pas limitatif) : le niveau limité par l'enveloppe de l'agent global qui contient l'organisation de plus haut niveau ou niveau macro. Dans cet environnement est construit le niveau principal des objets du système, celui qui contient les objets qu'on étudie, constitué du niveau méso. Ces objets sont eux-mêmes structurés par des objets dits terminaux, c'est-à-dire qui ne sont pas décomposables en objets plus élémentaires, c'est le niveau particulaire, ou niveau micro. Rien n'empêche d'ajouter d'autres niveaux si le problème le demande.

Nous ne négligeons pas pour autant les méthodes de conception de l'informatique, depuis les langages de conception comme UML ou SADT, jusqu'à l'algorithmique et aux structures de données informatiques. Si elles n'ont pas la puissance d'expression du discours mathématique à cause leur caractère schématique et simplificateur, elles ont en revanche une puissance élevée de communication par une explicitation directe[2]. Ceci est important dans un contexte pluridisciplinaire. De plus la formalisation conceptuelle permet de passer rapidement à la programmation informatique, pour répondre au deuxième objectif, celui de réaliser une plate-forme de simulation basée sur cette ontologie.

Ces deux démarches de formalisation, mathématique et informatique, ne sont pas antagonistes mais complémentaires et s'enrichissent de sens mutuellement lorsqu'elles renvoient l'une à l'autre la description d'un même objet de la réalité géographique. Notre méthode sera donc de présenter le plus souvent un objet ou un concept sous la forme d'une description dans le discours habituel du géographe, puis de le formaliser mathématiquement, et/ou de voir comment cela peut se traduire dans le cadre d'une formalisation informatique : conceptuelle, structurelle ou même algorithmique.

[2] au sens où il n'est pas nécessaire d'avoir une culture préalable du domaine, car ayant pris connaissance des règles syntaxiques très simple de ces langages semi-graphiques, tout est alors explicité sur un seul niveau, alors que les mathématiques, utilisent des définitions, des opérations et des théorèmes, qui relèvent d'une construction sémantique hiérarchisée, dont les niveaux inférieurs sont supposés acquis par le lecteur. Tout comme le discours philosophique, le discours mathématique est fortement implicite. Alors qu'en informatique, l'objectif final est toujours de programmer un ordinateur, auquel il faut tout expliciter dans les plus fins détails, car ce n'est qu'une machine. Néanmoins, c'est précisément le but de l'informatique que de donner à l'ordinateur une certaine « culture humaine » préalable, pour qu'on puisse communiquer avec lui de manière moins mécanique

Nous arrivons au second objectif (extérieur au programme COS-MA-GEM). Le travail théorique devra déboucher sur la conception et la réalisation d'une plate-forme évolutive et modulaire de modélisation/simulation orientée géographie, contenant un langage de spécification de modèles (plus précisément, plusieurs niveaux de langage, selon le niveau d'accès à la plate-forme). La classe des modèles capable d'être réifiés par ce langage sera conforme à cette ontologie formelle. Une fois la plate-forme construite, celle-ci devra permettre à un modélisateur géographe de spécifier un modèle par des outils interactifs simples, utilisant un langage à la fois graphique et à base de règles, de construire ou d'importer depuis un SIG la configuration initiale d'une scène géographique. Elle devra ensuite être capable de compiler le modèle et de lancer différentes simulations par le moteur de la plate-forme, pour extraire différentes observations permettant de tester des hypothèses et d'en vérifier la pertinence. A ce niveau (haut) d'accès pour la simulation, il doit exister plusieurs autres accès de plus bas niveau, permettant d'enrichir les fonctionnalités par l'adjonction de modules « métier » ou encore de faire évoluer le méta-modèle général et le langage, modifiant ainsi la classe des modèles susceptible d'être traitée.

P. Langlois, le 5 fev 07